JOURNAL OF NONPARAMETRIC STATISTICS
2022, VOL. 34, NO. 1, 22–57
https://doi.org/10.1080/10485252.2021.2015767

Taylor & Francis
Taylor & Francis Group

Check for updates

The hazard level set

Guangcai Mao [a], Yanyan Liu [b], Yuanshan Wu [c] and Qinglong Yang [c]

[a] School of Mathematics and Statistics, Central China Normal University, Wuhan, Hubei, People's Republic of China; [b] School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei, People's Republic of China; [c] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei, People's Republic of China

ABSTRACT

The hazard rate function plays a fundamental role in survival analysis. Its statistical inference methods have been systemically and exclusively studied. When does the hazard rate reach a particular warning level? This is a basic question of interest to the investigator but largely left to be explored in practice. We define a level set of hazard rate to address this issue and propose a kernel smoothing estimator for such a level set. In terms of the Hausdorff distance, we establish the consistency, convergence rate and asymptotic distribution of the level set estimator. The validity of the proposed confidence set, based on the bootstrap method, for the level set of hazard rate function is theoretically justified. We conduct comprehensive simulation studies to assess the finite-sample performance of the proposed method, which is further illustrated with a breast cancer study.

ARTICLE HISTORY
Received 1 May 2020
Accepted 28 November 2021

KEYWORDS
Bootstrap; empirical process; Hausdorff distance; random set; set inference

2010 MATHEMATICS SUBJECT CLASSIFICATIONS
62G05; 62H12; 62N02

1. Introduction

The hazard rate function, known as failure rate in reliability and survival analysis, and also as force of mortality in demographics, is a core concept in many fields. It specifies the instantaneous rate at which failures occur for subjects that are surviving at time t. For unit-variate survival time T, the associated hazard rate function is defined as

$$\lambda(t) = \lim_{s \to 0^+} \frac{\mathbb{P}(t \le T < t + s \mid T \ge t)}{s}.$$

The censoring, as a feature of survival data, makes the estimation for the hazard rate function $\lambda(t)$ nontrivial; there have been significant efforts to develop methodologies and theory for this challengeable issue. Watson and Leadbetter (1964) proposed a kernel smoothing estimator for the unit-variate hazard rate function, which was further furnished by Tanner and Wong (1983), Ramlau-Hansen (1983), Yandell (1983), and Lo, Mack, and Wang (1989), among others. Multivariate survival data, especially bivariate survival data, where each subject may experience several types of events or several subjects as a group experience one type of event, are commonly observed in modern biomedical study. As

CONTACT Yuanshan Wu wu@zuel.edu.cn School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, People's Republic of China

the bivariate hazard rate function is closely related with the bivariate survival function, its estimation method always stems from estimator for the bivariate survival function. In the framework of censoring, Burke (1988) proposed to estimate the bivariate survival function by suitably modifying the estimator proposed by Campbell and Földes (1982) to satisfy the monotonicity. Employing the product integral techniques and the Volterra integral equations theory, Dabrowska (1988, 1989) proposed a bivariate Kaplan–Meier estimate (Dabrowska representation). Pruitt (1991) proposed to estimate the bivariate survival function based on the self-consistency equations of the EM-algorithm. Prentice and Cai (1992) proposed a representation (Prentice–Cai representation) of bivariate survival function by using the marginal survival functions and the related covariance function. Furthermore, Gill, van der Laan, and Wellner (1995) studied the limiting distribution of the Dabrowska representation and the Prentice–Cai representation by applying the functional delta method.

The nonparametric estimation methods for the hazard rate function, univariate or bivariate, or their variety, have been extensively investigated. On the contrary, in practice we may be interested to know when the hazard rate reaches a particular warning level. In clinical study, the investigators are eager to understand at what time the hazard rate of tumor exceeds the prespecified warning level. It is also crucial to know when the hazard rate of default risk in the credit risk field and thereof ruin probability in insurance attain the warning levels that corporations can bear. Furthermore, regarding to such a time point patients or customers can be classified as time-dependent risk, higher or lower, groups. We define a level set of hazard rate, consisting of all the time points at which the hazard rate is equal to a given level. Utilising empirical process theory and technique of kernel smoothing, we further propose an estimator for such a level set and establish its consistency, convergence rate and asymptotic distribution in terms of the Hausdorff distance.

Our work is closely related with the set inference, which has been studied and promoted mainly in the society of econometrics. Gilstein and Leamer (1983) provided set consistent estimation in a class of likelihood models. Furthermore, Chernozhukov, Hong, and Tamer (2007) considered a general class of criterion-based econometric models which is identified on a set of parameters and proposed a valid method of obtaining confidence regions for such an identified set parameters by employing empirical process theory. Romano and Shaikh (2010) provided computationally intensive, yet feasible methods for inference in a very general class of partially identified econometric models. On the other hand, many efforts have been made to the inference for density level set. Under various metrics, the consistency and convergence rate for the density level set have been investigated by Polonik (1995), Tsybakov (1997), Walther (1997), Cuevas, González-Manteiga, and Rodríguez-Casal (2006), Rinaldo and Wasserman (2010). Furthermore, Singh, Scott, and Nowak (2009) established the minimaxity of the estimate of density level set and both Mammen and Polonik (2013) and Jankowski, Ji, and Stanberry (2014) proposed methods for constructing confidence sets for the density level sets by using variation of density function. Recently, Chen, Genovese, and Wasserman (2017), under the Hausdorff distance, have established the limiting distribution of the estimate of density level set and constructed the confidence set by using the bootstrap method.

Although there have been significant recent efforts to develop methodologies for various set inference in some particular fields, there is a paucity of methods with theoretical guarantees for set inference with censored survival data. From the perspective of the hazard rate function, which is a fundamental concept in survival analysis, we define the level set of the hazard rate and propose an estimator for it. In contrast to the density level set estimator proposed by Chen et al. (2017) where the kernel smoothing density function estimator naturally enjoys the independent and identically distributed (i.i.d.) expression, the non-i.i.d. expression of the kernel smoothing hazard rate function estimator imposes the challenging obstacle in theoretical derivations. We make much effort to surmount the difficulty by constructing the i.i.d. counterpart and further evaluating the approximating error, which is shown to be non-ignorable and further explicitly expressed via sample size and smoothing bandwidth. Based on these painstaking preparations and employing modern geometric techniques and empirical process theory, we rigorously establish the consistency, convergence rate and asymptotic distribution of the level set estimator. We further theoretically validate the bootstrap-based method to construct confidence set for the hazard level set. Numerical results show that the proposed method exhibits favourable performances in the finite-sample settings.

The rest of this paper is organised as follows. In Section 2, we propose the estimation method for the hazard level set. We establish the asymptotic properties of the proposed method in Section 3. In Section 4, we conduct simulation studies to evaluate its finite-sample performance and illustrate our method with application to a real data example. Some remarks are concluded in Section 5, and all the proofs are provided in Section 6.

2. Hazard level set

We first introduce some conventional notation in multivariate setting. For any $\boldsymbol{a} = (a_1, \dots, a_d)^{\mathrm{T}} \in \mathbb{R}^d$ and $\boldsymbol{b} = (b_1, \dots, b_d)^{\mathrm{T}} \in \mathbb{R}^d$, let $\min\{\boldsymbol{a}, \boldsymbol{b}\} = (\min\{a_1, b_1\}, \dots, \min\{a_d, b_d\})^{\mathrm{T}}$ and $I(\boldsymbol{a} \le \boldsymbol{b}) = \prod_{j=1}^{d} I(a_j \le b_j)$, where $I(\cdot)$ is the indicator function. Denote $\boldsymbol{a} \le \boldsymbol{b}$ if $a_j \le b_j$ for $j = 1, \dots, d$. Let $[\boldsymbol{a}, \boldsymbol{b}] = [a_1, b_1] \times \cdots \times [a_d, b_d]$ for $\boldsymbol{a} \le \boldsymbol{b}$. The Euclidean norm in $\mathbb{R}^d$ is denoted by $\|\cdot\|_2$. To present our work in a general framework, we consider d-variate survival time $\boldsymbol{T} = (T_1, \dots, T_d)^{\mathrm{T}}$ to denote the survival time for multiple events occurring on a subject or for multiple subjects in a cluster experiencing a single event. The d-variate hazard rate function is defined as

$$\lambda(\boldsymbol{t}) = \lim_{\boldsymbol{s} > \boldsymbol{0}, \|\boldsymbol{s}\|_2 \to 0} \frac{\mathbb{P}(\boldsymbol{t} \le \boldsymbol{T} < \boldsymbol{t} + \boldsymbol{s} \mid \boldsymbol{T} \ge \boldsymbol{t})}{s_1 \cdots s_d},$$

where $\boldsymbol{s} = (s_1, \dots, s_d)^{\mathrm{T}}$. Correspondingly, let $\boldsymbol{C} = (C_1, \dots, C_d)^{\mathrm{T}}$ denote the d-variate censoring time. Let $\boldsymbol{X} = \min\{\boldsymbol{T}, \boldsymbol{C}\}$ be the observed time and $\Delta = I(\boldsymbol{T} \le \boldsymbol{C})$ be the failure indicator. Suppose the study duration is on $[\boldsymbol{0}, \boldsymbol{\tau}_0]$, where $\boldsymbol{\tau}_0 = (\tau_{01}, \dots, \tau_{0d})^{\mathrm{T}}$ allowing different endpoints for each component of survival times. As usual, $\boldsymbol{T}$ and $\boldsymbol{C}$ are assumed to be independent.

For a specific $\ell > 0$, we define the level set of the hazard rate function λ as follows,

$$\mathcal{L} = \{\boldsymbol{t} \in [\boldsymbol{0}, \boldsymbol{\tau}_0] : \lambda(\boldsymbol{t}) = \ell\}.$$

Furthermore, we can define the upper hazard level set as $\{\boldsymbol{t} \in [\boldsymbol{0}, \boldsymbol{\tau}_0] : \lambda(\boldsymbol{t}) \ge \ell\}$ and the lower hazard level set as $\{\boldsymbol{t} \in [\boldsymbol{0}, \boldsymbol{\tau}_0] : \lambda(\boldsymbol{t}) \le \ell\}$, respectively. For $i = 1, \dots, n$, $(\boldsymbol{X}_i, \Delta_i)$ are assumed to be independent and identically distributed copies of $(\boldsymbol{X}, \Delta)$. Based on these observations, we ideally aim to estimate and infer these hazard level sets. First of all, we need to know how to estimate the hazard rate function $\lambda(\boldsymbol{t})$. Note that the cumulative hazard function can be deduced as

$$\Lambda(\boldsymbol{t}) = \int_{[\mathbf{0},\boldsymbol{t}]} \lambda(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}.$$

Let $H(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \le \boldsymbol{x})$ and $\bar{H}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} > \boldsymbol{x})$ denote the distribution function and survival function of $\boldsymbol{X}$, respectively. Based on the random censoring assumption, the cumulative hazard function can be rewritten as

$$\Lambda(\boldsymbol{t}) = \int_{[\mathbf{0},\boldsymbol{t}]} \frac{1}{\bar{H}(\boldsymbol{s}-)} H_{11}(\mathrm{d}\boldsymbol{s}),$$

where $H_{11}(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} \le \boldsymbol{x}, \Delta = 1)$. Plugging in the sample versions, we can estimate it by

$$\widehat{\Lambda}(\boldsymbol{t}) = \int_{[\mathbf{0},\boldsymbol{t}]} \frac{1}{\bar{H}_n(\boldsymbol{s})} H_{n11}(\mathrm{d}\boldsymbol{s}),$$

where $\bar{H}_n(\boldsymbol{x}) = n^{-1}\sum_{i=1}^{n} I(\boldsymbol{X}_i \ge \boldsymbol{x})$ and $H_{n11}(\boldsymbol{x}) = n^{-1}\sum_{i=1}^{n} I(\boldsymbol{X}_i \le \boldsymbol{x}) I(\Delta_i = 1)$. Utilising the kernel smoothing theory, the estimator of $\lambda(\boldsymbol{t})$ is given by

$$\widehat{\lambda}_n(\boldsymbol{t}) = \int K_n(\boldsymbol{t}-\boldsymbol{s})\widehat{\Lambda}(\mathrm{d}\boldsymbol{s}),$$

where $K_n(\boldsymbol{x}) = h_n^{-d} K(\boldsymbol{x}/h_n)$, $K(\cdot)$ is a d-variate kernel function, and h_n is bandwidth that could depend on the sample size n. Therefore, we can estimate the hazard level set $\mathcal{L}$ by

$$\widehat{\mathcal{L}}_n = \{\boldsymbol{t} \in [\mathbf{0}, \boldsymbol{\tau}_0] : \widehat{\lambda}_n(\boldsymbol{t}) = \ell\}.$$

If we consider the smoothed version of $\lambda(\boldsymbol{t})$, we have

$$\lambda_n(\boldsymbol{t}) = \int K_n(\boldsymbol{t}-\boldsymbol{s})\lambda(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \int K_n(\boldsymbol{t}-\boldsymbol{s})\Lambda(\mathrm{d}\boldsymbol{s}).$$

Correspondingly, the hazard level set can be also given by

$$\mathcal{L}_n = \{\boldsymbol{t} \in [\mathbf{0}, \boldsymbol{\tau}_0] : \lambda_n(\boldsymbol{t}) = \ell\}.$$

Motivated by Chen et al. (2017), we really discuss in our work relations between the hazard level sets $\mathcal{L}_n$ and $\widehat{\mathcal{L}}_n$ from perspectives of theoretical derivations and practical applications. Specifically, while the bias of $\lambda_n(\boldsymbol{t}) - \lambda(\boldsymbol{t})$ may be analysed theoretically, in practice, it cannot be accurately estimated. On the contrary, as an unbiased estimator, $\widehat{\lambda}_n(\boldsymbol{t})$ converges to $\lambda_n(\boldsymbol{t})$ at a much faster rate. As a result, $\lambda_n(\boldsymbol{t})$, preserving the salient structure of $\lambda(\boldsymbol{t})$, could be viewed as an alternative estimand.

It is conventional to use the Hausdorff distance to measure how far two subsets of a metric space are from each other. Specifically, for any two subsets A and B of metric space $(\mathbb{R}^d, \rho)$, the Hausdorff distance between A and B is defined by

$$\begin{aligned}\mathsf{Haus}(A,B) &= \inf\{\epsilon > 0 : A \subset B \oplus \epsilon \text{ and } B \subset A \oplus \epsilon\}\\ &= \max\left\{\sup_{\boldsymbol{x}\in B} \rho(\boldsymbol{x}, A), \sup_{\boldsymbol{x}\in A} \rho(\boldsymbol{x}, B)\right\},\end{aligned}$$

where ρ is the Euclidean distance, $A \oplus \epsilon = \bigcup_{\boldsymbol{x}\in A}\{\boldsymbol{y} : \rho(\boldsymbol{y}, \boldsymbol{x}) \le \epsilon\}$, and $\rho(\boldsymbol{x}, A) = \inf\{\rho(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{y} \in A\}$. We will use the Hausdorff distance to evaluate the distance between the hazard level sets.

3. Asymptotic theory

Let $\boldsymbol{BC}^k(\mathcal{D})$ denote the collection of functions defined on $\mathcal{D}$ with bounded and continuous kth-order partial derivatives. We impose the following assumptions throughout our discussions.

(i) $\overline{H}(\boldsymbol{\tau}_0-) > 0$.
(ii) The symmetric kernel function $K(\cdot)$ vanishes on the outside of $[\mathbf{-1}, \mathbf{1}]$ and has continuous third-order partial derivatives.
(iii) $|\log h_n|/\log\log n \to \infty$ and $nh_n^{d+2}/|\log h_n| \to \infty$; h_n monotonically converges to zero and nh_n^d monotonically converges to infinity as $n \to \infty$.
(iv) The hazard rate function λ is bounded over $[\mathbf{0}, \boldsymbol{\tau}_0]$. There exists some positive constant c_0 such that $\|\nabla\lambda_n(\boldsymbol{t})\|_2 > c_0$ for all $\boldsymbol{t} \in \mathcal{L}_n$.
(v) $\mathcal{L}_n$ is a manifold lies in $[\mathbf{0}, \boldsymbol{\tau}]$ for some $\boldsymbol{\tau} < \boldsymbol{\tau}_0$.

Assumption (i) is imposed to make the cumulative hazard function to be well-defined, which is conventional in survival analysis. For ease of exposition, in assumption (ii), the support of the kernel is constricted on $[\mathbf{-1}, \mathbf{1}]$ instead of an arbitrary compact set. Such an assumption is also imposed in Ramlau-Hansen (1983), Diehl and Stute (1988), Giné and Guillou (2001), and Calonico, Cattaneo, and Farrell (2017). Assumption (iii) states that the bandwidth converges to zero at certain rate regarding to the sample size. Assumption (iv) excludes the situation in which the hazard rate function λ is horizontal, which is also imposed in Mammen and Polonik (2013), Laloë and Servien (2013), and Chen et al. (2017) for set inference. Assumption (v) is imposed to avoid the unstability of the hazard rate function (level set) estimation near the end time of the study. For a multivariate function $f : \mathbb{R}^d \to \mathbb{R}$ and $f \in \boldsymbol{BC}^2([\mathbf{0}, \boldsymbol{\tau}])$, we define the quantities

$$\|f\|_{0,\max} = \sup_{\boldsymbol{x}\in[\mathbf{0},\boldsymbol{\tau}]} |f(\boldsymbol{x})|,$$

$$\|f\|_{1,\max} = \sup_{\boldsymbol{x}\in[\mathbf{0},\boldsymbol{\tau}]} \max_{1\le i\le d} \left|\frac{\partial f(\boldsymbol{x})}{\partial x_i}\right|,$$

$$\|f\|_{2,\max} = \sup_{\boldsymbol{x}\in[\mathbf{0},\boldsymbol{\tau}]} \max_{1\le i,j\le d} \left|\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}\right|,$$

$$\|f\|^*_{k,\max} = \max\{\|f\|_{j,\max} : j = 0, \ldots, k\}$$

for $k = 0, 1, 2$. The following theorem shows the convergence rate between the estimated hazard rate function and the smoothed counterpart.

Theorem 3.1: *Suppose that assumptions* (i)–(iv) *hold. Then for $k = 0, 1, 2$, we have*

$$\|\widehat{\lambda}_n - \lambda_n\|^*_{k,\max} = O\left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2k}}}\right).$$

Actually, we establish the almost sure (*a.s.*) convergence rate but we suppress the abbreviation *a.s.* hereafter if there is no confusion. Whenever two smoothed hazard rate functions are sufficiently closed to each other, it could be expected that the corresponding hazard level sets are closed in terms of the Hausdorff distance, which is summarised in the following theorem.

Theorem 3.2: *Suppose that assumptions* (i)–(v) *hold. Then we have*

$$\mathit{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) = O(\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}) = O\left(\sqrt{\frac{|\log h_n|}{nh_n^d}}\right).$$

We establish the convergence rate for the almost sure convergence, which is used to build the following theorem.

Theorem 3.3: *Suppose that assumptions* (i)–(v) *hold. Then we have*

$$\sup_{x \in \mathcal{L}_n} \left| \frac{|\mathbb{G}_n(f_x)| - \sqrt{nh_n^d}\rho(x, \widehat{\mathcal{L}}_n)}{\sqrt{nh_n^d}\rho(x, \widehat{\mathcal{L}}_n)} \right| = O(\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max}) = O\left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2}}}\right),$$

where

$$\mathbb{G}_n(f) = \int_{[0,\tau_0]} f(s)\,\mathrm{d}[n^{1/2}\{\widehat{\Lambda}(s) - \Lambda(s)\}]$$

and f_x is defined as

$$f_x(s) = \frac{1}{\sqrt{h_n^d}\|\nabla\lambda_n(x)\|_2} K\left(\frac{x - s}{h_n}\right), \quad x \in \mathcal{L}_n.$$

Theorem 3.3 shows that the projection distance onto the level set $\widehat{\mathcal{L}}_n$ can be approximated well at certain rate by a functional empirical process. Thus, we collect the functions together by defining

$$\mathscr{F} = \left\{ f_x(\cdot) = \frac{1}{\sqrt{h_n^d}\|\nabla\lambda_n(x)\|_2} K\left(\frac{x - \cdot}{h_n}\right) : x \in \mathcal{L}_n \right\}.$$

Let $\mathbb{B}$ be the Gaussian process indexed by $\mathscr{F}$ such that for any $f_1 \in \mathscr{F}$ and $f_2 \in \mathscr{F}$,

$$\mathbb{B}(f_1) \stackrel{D}{=} N(0, \mathbb{E}\{f_1^2(X)I(\Delta = 1)\bar{H}^{-2}(X-)\})$$

and

$$\mathrm{Cov}(\mathbb{B}(f_1), \mathbb{B}(f_2)) = \mathbb{E}\{f_1(X)f_2(X)I(\Delta = 1)\bar{H}^{-2}(X-)\},$$

where $\stackrel{D}{=}$ stands for the equal in the sense of distribution.

Theorem 3.4: *Suppose that assumptions* (i)–(v) *hold. If $nh_n^{5d}(\log\log n)^4/\log^3 n \to 0$ as $n \to \infty$, then we have*

$$\sup_t \left| \mathbb{P}(\sqrt{nh_n^d}\,\mathrm{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \le t) - \mathbb{P}\left(\sup_{f \in \mathscr{F}} |\mathbb{B}(f)| \le t\right) \right|$$

$$= O\left(\left(\frac{\log^7 n}{nh_n^d}\right)^{1/8} + \left(\frac{\log^3 n}{nh_n^{d+2}}\right)^{1/2}\right).$$

Theorem 3.4 demonstrates that the distribution of the Hausdorff distance between $\widehat{\mathcal{L}}_n$ and $\mathcal{L}_n$ can be approximated by that of the supremum functional Gaussian process. However, we can not directly use this theorem to construct a confidence set for $\mathcal{L}_n$ as the limiting distribution is too complicated to be applied. We employ the bootstrap method to construct the confidence set. For ease of exposition, denote $W_n = \mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n)$ and let $w_{1-\alpha}$ be the $1-\alpha$ quantile of the distribution function of W_n for a given $\alpha \in (0, 1)$. Obviously, it holds that

$$\mathbb{P}(\mathcal{L}_n \subset \widehat{\mathcal{L}}_n \oplus w_{1-\alpha}) \geq 1 - \alpha.$$

Let $\mathcal{O}_i = (\boldsymbol{X}_i, \Delta_i)$ be the ith observation, $i = 1, \ldots, n$. Sampling n times from these observations with replacement, we obtain the bootstrap samples $\mathcal{O}_1^*, \ldots, \mathcal{O}_n^*$. Thus, we can construct the hazard level set $\widehat{\mathcal{L}}_n^*$ based on the bootstrapped samples. Let $W_n^* = \mathsf{Haus}(\widehat{\mathcal{L}}_n^*, \widehat{\mathcal{L}}_n)$ and $w_{1-\alpha}^*$ be the $1-\alpha$ quantile of the distribution function of W_n^*. Next theorem shows that the confidence set for $\mathcal{L}_n$ constructed by the bootstrap method is asymptotically valid.

Theorem 3.5: *Suppose that assumptions of Theorem* 3.4 *hold. Then there exists* $\mathscr{K}_n$ *such that*

$$\sup_t |\mathbb{P}(\sqrt{nh_n^d}\, \mathit{Haus}(\widehat{\mathcal{L}}_n^*, \widehat{\mathcal{L}}_n) \leq t | \mathcal{O}_1, \ldots, \mathcal{O}_n) - \mathbb{P}(\sqrt{nh_n^d}\, \mathrm{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \leq t)|$$
$$= O\left(\left(\frac{\log^7 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right)$$

for all $(\mathcal{O}_1, \ldots, \mathcal{O}_n) \in \mathscr{K}_n$ *and* $\mathbb{P}(\mathscr{K}_n) \geq 1 - c_1 \exp(-c_2 nh_n^{d+2})$ *for some constants* c_1 *and* c_2. *Thus, we have*

$$\mathbb{P}(\mathcal{L}_n \subset \widehat{\mathcal{L}}_n \oplus w_{1-\alpha}^*) \geq 1 - \alpha + O\left(\left(\frac{\log^7 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right).$$

In practice, repeating the sampling procedure N times leads to N realisations of W_n^*; the $1-\alpha$ quantile among these N realisations, denoted by $\widehat{w}_{1-\alpha}^*$, can be used to estimate $w_{1-\alpha}^*$. Thus, the $1-\alpha$ confidence set for $\mathcal{L}_n$ is given by $\widehat{\mathcal{L}}_n \oplus \widehat{w}_{1-\alpha}^*$.

Remark 3.1: $nh_n^{d+2}/|\log h_n| \to \infty$ in assumption (iii) implies that $nh_n^{d+2} \to \infty$, which leads to $\log n - (d+2)|\log h_n| > 0$. Thus, we have $|\log h_n| = O(\log n)$. On the other hand, it follows from $|\log h_n|/\log\log n \to \infty$ stated in assumption (iii) that $h_n \log^a n \to 0$ for any $a > 0$. Combining with the extra assumption $nh_n^{5d}(\log\log n)^4/\log^3 n \to 0$ in Theorems 3.4 and 3.5, we have $nh_n^{5d+1} \to 0$, which implies that $\log n = O(|\log h_n|)$. Hence, we conclude that

$$|\log h_n| = O(\log n), \quad \log n = O(|\log h_n|),$$

which will be used repeatedly in proofs of Theorems 3.4 and 3.5.

Remark 3.2: If we further assume that $nh_n^{d+2}/\log^7 n \to \infty$, which is a strengthened version of $nh_n^{d+2}/|\log h_n| \to \infty$ based on Remark 3.1, then the established rates in Theorems 3.4 and 3.5 both converge to zero. As a result, we can theoretically justify the weak convergence of the Haudorff distance between hazard level sets and validate its bootstrapped approximation. Actually, for any $d+2 < \nu_0 \le 5d$, let h_n be the order of n^{-1/ν_0}, then assumption $nh_n^{d+2}/\log^7 n \to \infty$ and the other assumptions in Theorems 3.4 and 3.5 are all satisfied.

Now we have systemically established the asymptotic properties of the estimate of the hazard level set $\mathcal{L}_n$, which backups the practical applications of our proposed method. The proofs of theorems are deferred to Section 6. Furthermore, we can also establish the asymptotic properties of the estimators of the upper and lower hazard level sets without much more efforts. We skip the related discussions to keep our focus on the hazard level set only.

The bandwidth choice is crucial for kernel smoothing method. Intuitively, bias between λ and λ_n is decreasing and that between λ_n and $\widehat{\lambda}_n$ is increasing when the bandwidth is decreasing. We employ the cross validation score function (Patil 1993) to choose the optimal bandwidth. Specifically, we define the cross validation criterion

$$\mathrm{CV}_n(h_n) = \int_{[\mathbf{0},\boldsymbol{\tau}_0]} \widehat{\lambda}_n^2(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} - 2\sum_{i=1}^{n} \frac{\widehat{\lambda}_n^{(-i)}(\boldsymbol{X}_i) I(\Delta_i = 1)}{\sum_{j=1}^{n} I(\boldsymbol{X}_j \ge \boldsymbol{X}_i)}, \quad (1)$$

where $\widehat{\lambda}_n^{(-i)}$ is the leave-one-out version of $\widehat{\lambda}_n$ by deleting the ith observation. The optimal bandwidth is defined by $\widehat{h}_n = \arg\min_{h_n} \mathrm{CV}_n(h_n)$, which is recommended in our numerical studies.

4. Numerical studies

We conducted simulation studies to evaluate the finite-sample performance of the proposed method. For ease of exposition, we considered bivariate survival time, i.e. $d = 2$, in numerical studies.

Example 4.1: We first generated the survival time $\boldsymbol{T} = (T_1, T_2)^{\mathrm{T}}$ from the bivariate hazard rate function

$$\lambda(t_1, t_2) = [\exp\{-(t_1-2)^2\} + 0.1] \times [\exp\{-(t_2-2)^2\} + 0.1]$$

and specified the hazard level $\ell = (110 + 11e)/(100e)$. We considered the kernel function $K(t_1, t_2) = 0.16 \times \exp\{-(t_1^2 + t_2^2)/2\} I(-3 \le t_1 \le 3) I(-3 \le t_2 \le 3)$. Note that T_1 and T_2 were independently generated. We set the censoring time $\boldsymbol{C} = \min\{\widetilde{\boldsymbol{C}}, \boldsymbol{\tau}\}$, where $\widetilde{\boldsymbol{C}}$ was generated from the uniform distribution on $[\mathbf{0}, \boldsymbol{\tau} + \boldsymbol{L}]$. The censoring-tuning parameter $\boldsymbol{L}$ and study duration $\boldsymbol{\tau}$ were chosen to yield the censoring rates of (20%, 20%), (50%, 20%) and (50%, 50%). We considered sample sizes $n = 100, 200$ and 400. The smoothing parameter was selected by minimising (1). For each configuration, we repeated 500 simulations, and for each replicated data set 200 bootstrap samples were generated for constructing the confidence set.

Table 1. Simulation results for the proposed hazard level set estimate in Example 4.1.

CR	n	Median	Mean	CP
(20%, 20%)	100	0.577	0.696	0.935
	200	0.510	0.636	0.954
	400	0.368	0.505	0.960
(50%, 20%)	100	0.709	0.800	0.914
	200	0.586	0.674	0.945
	400	0.498	0.596	0.941
(50%, 50%)	100	0.707	0.750	0.840
	200	0.659	0.695	0.904
	400	0.576	0.599	0.927

Notes: CR, the censoring rate; Median, the median of the Hausdorff distance among 500 simulations; Mean, the mean of the Hausdorff distance among 500 simulations; and CP, the coverage probability of 95% bootstrapped confidence sets.

Table 1 summarises simulation results under different sample sizes and censoring rates. As expected, both the median and the mean of the Hausdorff distances among 500 simulations decrease dramatically when the sample size is increased from 100 to 400. Furthermore, the coverage probabilities of 95% bootstrapped confidence sets are enhanced to be around the nominal level. As usual, the performance of the proposed method could be negatively affected by the higher censoring rates. However, it is greatly improved when n is increased to 400. As a result, our method exhibits favourable performances in the finite-sample settings. To gain more insight of the proposed bootstrap method for constructing the confidence set, based on one simulated dataset with $n = 400$ and the censoring rate (50%, 20%), we plot in the right panel of Figure 1 the hazard level set $\mathcal{L}_n$ and the estimated hazard level set $\widehat{\mathcal{L}}_n$ as well as the 95% bootstrapped confidence set for $\mathcal{L}_n$. It can be seen that the proposed confidence set covers the hazard level set $\mathcal{L}_n$ very well. Correspondingly, the optimal bandwidth selection by minimising (1) is illustrated in the left panel of Figure 1, showing that the cross validation criterion is feasible.

Example 4.2: We further considered a situation where T_1 and T_2 were generated dependently from the hazard function $\lambda(t_1, t_2) = f(t_1, t_2)/S(t_1, t_2)$, where

$$f(t_1, t_2) = 0.336\{(t_1 - 2)^2 + (t_2 - 2)^2 + 1\}^{-3/4} \exp[-\{(t_1 - 2)^2 + (t_2 - 2)^2 + 1\}^{1/4}] \times I(0 < t_1 < 20)I(0 < t_2 < 20)$$

and

$$S(t_1, t_2) = \int_{t_1}^{\infty} \int_{t_2}^{\infty} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The hazard level was fixed at $\ell = 0.2$. We kept the remaining setup the same as Example 4.1. The simulation results are summarised in Table 2 and the evaluations of the optimal bandwidth selection procedure and the proposed bootstrapped confidence set method are presented in Figure 2, from which we can draw similar conclusions.

Example 4.3: As a real example, we now apply the proposed method to the German Breast Cancer study. The data is available from http://www.umass.edu/statdata/statdata/data. In

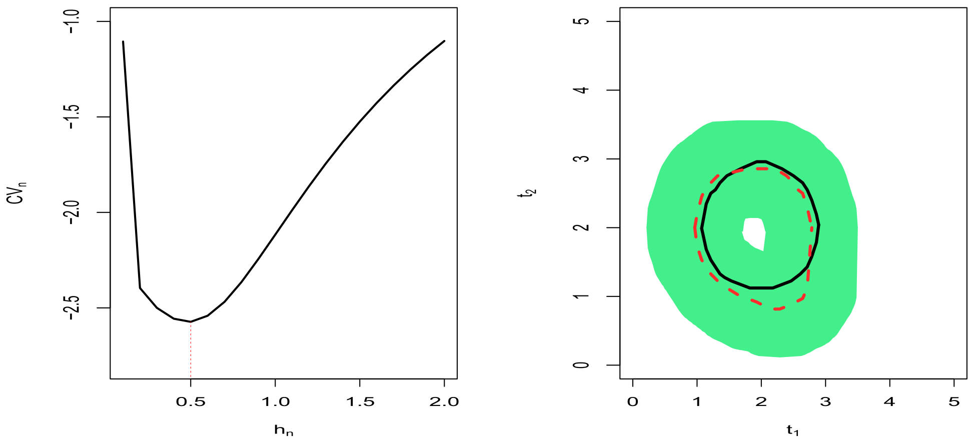

Figure 1. The left panel shows the optimal bandwidth selection based on one simulated dataset with $n = 400$ and the censoring rate (50%, 20%) in Example 4.1. The right panel correspondingly shows the hazard level set $\mathcal{L}_n$ (solid lines), the estimated hazard level set $\widehat{\mathcal{L}}_n$ (dotted lines), and the 95% bootstrapped confidence set for $\mathcal{L}_n$.

Table 2. Simulation results for the proposed hazard level set estimate in Example 4.2.

CR	n	Median	Mean	CP
(20%, 20%)	100	0.872	0.956	0.936
	200	0.718	0.792	0.962
	400	0.577	0.634	0.962
(50%, 20%)	100	1.104	1.246	0.886
	200	0.918	1.077	0.914
	400	0.777	0.897	0.944
(50%, 50%)	100	1.327	1.386	0.789
	200	1.122	1.274	0.896
	400	0.878	1.053	0.921

Notes: CR, the censoring rate; Median, the median of the Hausdorff distance among 500 simulations; Mean, the mean of the Hausdorff distance among 500 simulations; and CP, the coverage probability of 95% bootstrapped confidence sets.

the study, a total of 686 patients with primary node positive breast cancer were recruited between July 1984 and December 1989. The time to the recurrence of breast cancer and the time to death were monitored during study along with the censoring rates 56.4% and 75.1%, respectively. To provide an overall view, we plot in Figure 3 the heat map for the hazard rate function estimate. It reveals that both lower-left and top-right regions demonstrated very higher hazard rates. Furthermore, the patients in the study tended to suffer from much more serious life-threatening disease than the tumor recurrence. Figure 4 shows the optimal bandwidth selection procedure as well as the hazard level set estimate with $\ell = 0.002$ and the 95% confidence set based on 200 bootstrapped samples. In general, it captures the trend of the time set in which the hazard rate function reached the given level 0.002, in comparison with Figure 3.

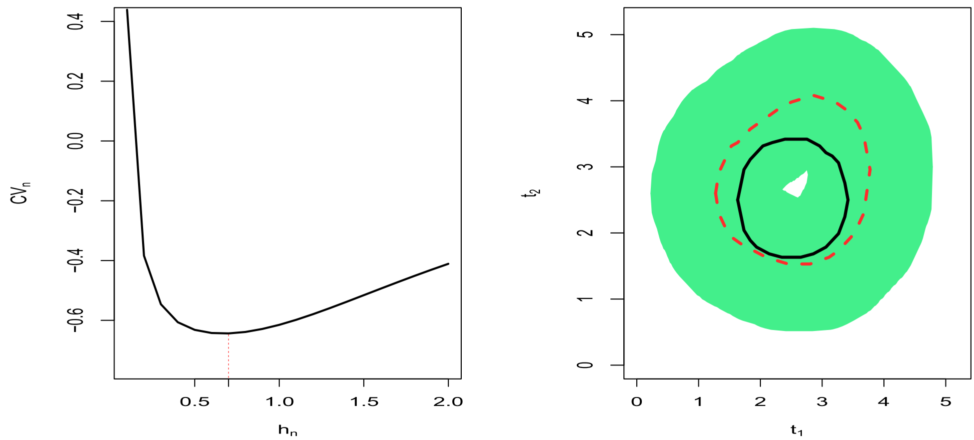

Figure 2. The left panel shows the optimal bandwidth selection based on one simulated dataset with $n = 400$ and the censoring rate (50%, 20%) in Example 4.2. The right panel correspondingly shows the hazard level set $\mathcal{L}_n$ (solid lines), the estimated hazard level set $\widehat{\mathcal{L}}_n$ (dotted lines), and the 95% bootstrapped confidence set for $\mathcal{L}_n$.

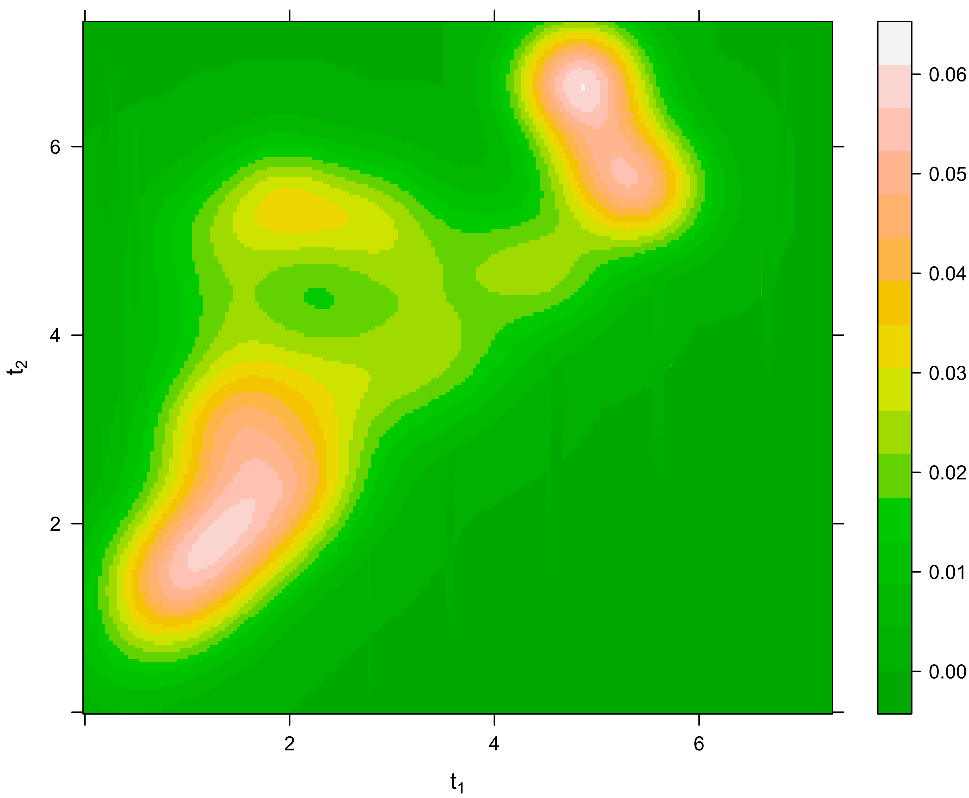

Figure 3. The heat map of the estimated hazard rate function in the German Breast Cancer study.

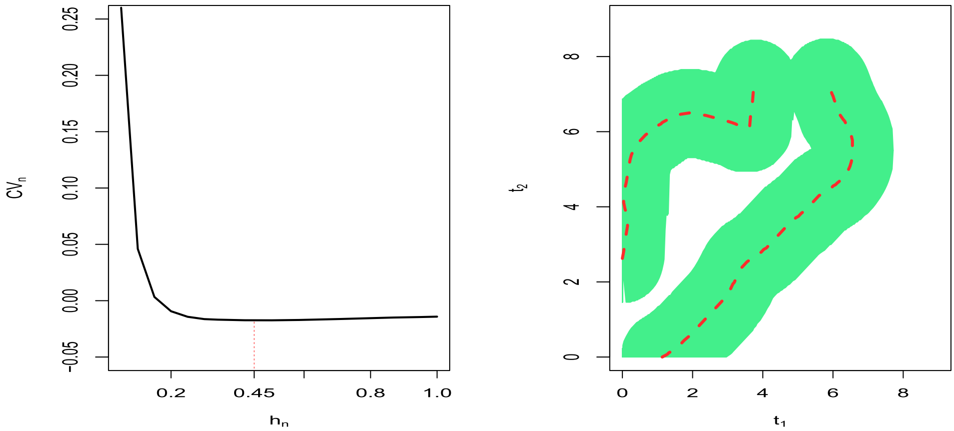

Figure 4. The left panel shows the optimal bandwidth selection for the German Breast Cancer study. The right panel correspondingly shows the hazard level set estimate with $\ell = 0.002$ (dotted lines) and the 95% bootstrapped confidence set.

5. Concluding remarks

The hazard rate function is the core concept in survival analysis. There has been a long history of researches on its nonparametric estimation methods. From a novel perspective, we define the hazard level set and comprehensively study its theoretical properties, including the convergence rate and weak convergence under the Hausdorff distance, and further propose a valid bootstrap method to construct the confidence set. Numerical results demonstrate that the proposed method performs favourably in finite-sample settings.

In clinical study, the investigators are often interested to know at what time the hazard rate of some tumor exceeds the prespecified warning level. In finance, it is also vital to understand when the hazard rate of default risk achieves the warning level that corporations can bear. The related hazard level could be considered a guidance for what diagnostic strategy be further adopted. Our work endeavours to initiate theoretical basis and illuminate potential application of the hazard level set in survival analysis.

6. Theoretical proofs

We provide proofs of theorems presented in Section 3. For ease of exposition, we consider the case $d = 2$ while our derivations can be extended to higher dimension case, mainly involving the complicated multivariate integral transformation.

Proof of Theorem 3.1: We first prove that it holds for $k = 0$. For $\boldsymbol{t} = (t_1, t_2)^{\mathrm{T}} \in [\boldsymbol{0}, \boldsymbol{\tau}]$, under assumption (ii), the expressions of $\widehat{\lambda}_n$ and λ_n can be further refined as

$$\widehat{\lambda}_n(\boldsymbol{t}) = \int_{\mathcal{B}_n(\boldsymbol{t})} K_n(\boldsymbol{t} - \boldsymbol{s})\widehat{\Lambda}(\mathrm{d}\boldsymbol{s}) \quad \text{and} \quad \lambda_n(\boldsymbol{t}) = \int_{\mathcal{B}_n(\boldsymbol{t})} K_n(\boldsymbol{t} - \boldsymbol{s})\Lambda(\mathrm{d}\boldsymbol{s}),$$

where $\mathcal{B}_n(\boldsymbol{t}) = [t_1 - h_n, t_1 + h_n] \times [t_2 - h_n, t_2 + h_n]$. Then we have

$$\widehat{\lambda}_n(\boldsymbol{t}) - \lambda_n(\boldsymbol{t}) = \int_{\mathcal{B}_n(\boldsymbol{t})} K_n(\boldsymbol{t} - \boldsymbol{s})(\widehat{\Lambda} - \Lambda)(\mathrm{d}\boldsymbol{s}).$$

Under assumption (ii), using integration by parts and noting that $K_n(\boldsymbol{t}-\boldsymbol{s})$ and $\nabla K_n(\boldsymbol{t}-\boldsymbol{s})$, as functions of $\boldsymbol{s}$, are zero at the boundary of $\mathcal{B}_n(\boldsymbol{t})$, we further have

$$\widehat{\lambda}_n(\boldsymbol{t})-\lambda_n(\boldsymbol{t})=\int_{\mathcal{B}_n(\boldsymbol{t})}(\widehat{\Lambda}-\Lambda)(\boldsymbol{s})\,\mathrm{d}K_n(\boldsymbol{t}-\boldsymbol{s}). \tag{2}$$

Rewrite

$$\begin{aligned}\widehat{\Lambda}(\boldsymbol{t})-\Lambda(\boldsymbol{t})&=\int_{[\boldsymbol{0},\boldsymbol{t}]}\frac{1}{\bar{H}_n(\boldsymbol{s})}H_{n11}(\mathrm{d}\boldsymbol{s})-\int_{[\boldsymbol{0},\boldsymbol{t}]}\frac{1}{\bar{H}(\boldsymbol{s}-)}H_{11}(\mathrm{d}\boldsymbol{s})\\&=\int_{[\boldsymbol{0},\boldsymbol{t}]}\left\{\frac{1}{\bar{H}_n(\boldsymbol{s})}-\frac{1}{\bar{H}(\boldsymbol{s}-)}\right\}H_{n11}(\mathrm{d}\boldsymbol{s})+\int_{[\boldsymbol{0},\boldsymbol{t}]}\frac{1}{\bar{H}(\boldsymbol{s}-)}(H_{n11}-H_{11})(\mathrm{d}\boldsymbol{s})\\&\equiv\xi_{n1}(\boldsymbol{t})+\xi_{n2}(\boldsymbol{t}),\end{aligned}$$

where $\xi_{n1}(\boldsymbol{t})$ and $\xi_{n2}(\boldsymbol{t})$ are self-explained from the expression. As a result, (2) can be further deduced as

$$\begin{aligned}\widehat{\lambda}_n(\boldsymbol{t})-\lambda_n(\boldsymbol{t})&=\int_{\mathcal{B}_n(\boldsymbol{t})}\xi_{n1}(\boldsymbol{s})\,\mathrm{d}K_n(\boldsymbol{t}-\boldsymbol{s})+\int_{\mathcal{B}_n(\boldsymbol{t})}\xi_{n2}(\boldsymbol{s})\,\mathrm{d}K_n(\boldsymbol{t}-\boldsymbol{s})\\&=\eta_{n1}(\boldsymbol{t})+\eta_{n2}(\boldsymbol{t}),\end{aligned} \tag{3}$$

where the definitions of $\eta_{n1}(\boldsymbol{t})$ and $\eta_{n2}(\boldsymbol{t})$ are also self-explained. We first consider the term $\eta_{n1}(\boldsymbol{t})$. Let $(t_1-s_1)/h_n=u$ and $(t_2-s_2)/h_n=v$, then using assumption (ii), we have

$$\begin{aligned}\eta_{n1}(\boldsymbol{t})=\eta_{n1}(t_1,t_2)&=h_n^{-2}\int_{-1}^{1}\int_{-1}^{1}\xi_{n1}(t_1-h_nu,t_2-h_nv)K(\mathrm{d}u,\mathrm{d}v)\\&=h_n^{-2}\int_{-1}^{1}\int_{-1}^{1}\xi_{n1}(t_1+h_nu,t_2+h_nv)K(\mathrm{d}u,\mathrm{d}v).\end{aligned} \tag{4}$$

Furthermore, the integrals of $\int_{-1}^{1}\int_{-1}^{1}\xi_{n1}(t_1+h_nu,t_2)K(\mathrm{d}u,\mathrm{d}v)$, $\int_{-1}^{1}\int_{-1}^{1}\xi_{n1}(t_1,t_2)K(\mathrm{d}u,\mathrm{d}v)$, and $\int_{-1}^{1}\int_{-1}^{1}\xi_{n1}(t_1,t_2+h_nv)K(\mathrm{d}u,\mathrm{d}v)$ are all zero based on assumption (ii). Hence, (4) can be written as

$$\begin{aligned}\eta_{n1}(\boldsymbol{t})=h_n^{-2}&\int_{-1}^{1}\int_{-1}^{1}\{\xi_{n1}(t_1+h_nu,t_2+h_nv)-\xi_{n1}(t_1+h_nu,t_2)\\&-\xi_{n1}(t_1,t_2+h_nv)+\xi_{n1}(t_1,t_2)\}K(\mathrm{d}u,\mathrm{d}v)\\&=h_n^{-2}\int_{-1}^{1}\int_{-1}^{1}\left\{\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\xi_{n1}(\mathrm{d}s_1,\mathrm{d}s_2)\right\}K(\mathrm{d}u,\mathrm{d}v)\\&=h_n^{-2}\int_{-1}^{1}\int_{-1}^{1}\left\{\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{\bar{H}(s_1-,s_2-)-\bar{H}_n(s_1,s_2)}{\bar{H}(s_1-,s_2-)\bar{H}_n(s_1,s_2)}\right.\\&\left.\times H_{n11}(\mathrm{d}s_1,\mathrm{d}s_2)\right\}K(\mathrm{d}u,\mathrm{d}v).\end{aligned} \tag{5}$$

Note that h_n monotonically converges to zero as $n\to\infty$. Thus, there exists some $\boldsymbol{\tau}^*<\boldsymbol{\tau}_0$ such that $\tau_1+h_n<\tau_1^*$ and $\tau_2+h_n<\tau_2^*$ when n is sufficiently large. Then, under assumption (i), we have

$$\sup_{t_1,t_2,u,v}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{\bar H(s_1-,s_2-)-\bar H_n(s_1,s_2)}{\bar H(s_1-,s_2-)\bar H_n(s_1,s_2)}H_{n11}(\mathrm{d}s_1,\mathrm{d}s_2)\right|$$
$$\le c_3\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}|\bar H_n(\boldsymbol{s})-\bar H(\boldsymbol{s}-)|\times\sup_{t_1,t_2,u,v}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}H_{n11}(\mathrm{d}s_1,\mathrm{d}s_2)\right|, \quad (6)$$

where c_3 is some positive constant. Furthermore

$$\sup_{t_1,t_2,u,v}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}H_{n11}(\mathrm{d}s_1,\mathrm{d}s_2)\right|$$
$$\le\sup_{t_1,t_2,u,v}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}(H_{n11}-H_{11})(\mathrm{d}s_1,\mathrm{d}s_2)\right|$$
$$+\sup_{t_1,t_2,u,v}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}H_{11}(\mathrm{d}s_1,\mathrm{d}s_2)\right|$$
$$\le 4h_n^{-2}\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}|H_{n11}(\boldsymbol{s})-H_{11}(\boldsymbol{s})|+\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}|H''_{11}(\boldsymbol{s})|, \quad (7)$$

where H''_{11} is the sub-density function corresponding to the sub-distribution function H_{11}. It follows from the iterated logarithm that

$$\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}|\bar H_n(\boldsymbol{s})-\bar H(\boldsymbol{s}-)|=O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad (8)$$

and

$$\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}|H_{n11}(\boldsymbol{s})-H_{11}(\boldsymbol{s})|=O\left(\sqrt{\frac{\log\log n}{n}}\right). \quad (9)$$

Using the boundedness condition of λ in assumption (iv), we have, for some constant c_4,

$$\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}|H''_{11}(\boldsymbol{s})|=\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}\bar H(\boldsymbol{s})\lambda(\boldsymbol{s})\le\sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]}\lambda(\boldsymbol{s})\le c_4. \quad (10)$$

The total variation norm of K over $[-\mathbf{1},\mathbf{1}]$ is denoted by $\|K\|_{\mathrm{TV}}$, which is bounded by some constant c_5 under assumption (ii). Therefore, combining (5), (6), (7), (8), (9), and (10), we have

$$\sup_{\boldsymbol{t}}|\eta_{n1}(\boldsymbol{t})|$$
$$\le\sup_{t_1,t_2,u,v}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{\bar H(s_1-,s_2-)-\bar H_n(s_1,s_2)}{\bar H(s_1-,s_2-)\bar H_n(s_1,s_2)}H_{n11}(\mathrm{d}s_1,\mathrm{d}s_2)\right|\cdot\|K\|_{\mathrm{TV}}$$

$$\leq c_3 O\left(\sqrt{\frac{\log\log n}{n}}\right)\left(4h_n^{-2} O\left(\sqrt{\frac{\log\log n}{n}}\right) + c_4\right) c_5$$
$$= O\left(\frac{\log\log n}{nh_n^2} + \sqrt{\frac{\log\log n}{n}}\right). \tag{11}$$

Using the similar arguments, we rewrite $\eta_{n2}(\boldsymbol{t})$ as

$$\eta_{n2}(\boldsymbol{t}) = \int_{-1}^{1}\int_{-1}^{1}\left\{h_n^{-2}\int_{t_1}^{t_1+h_n u}\int_{t_2}^{t_2+h_n v}\frac{1}{\bar{H}(s_1-,s_2-)}\right.$$
$$\left.\times (H_{n11} - H_{11})(\mathrm{d}s_1, \mathrm{d}s_2)\right\} K(\mathrm{d}u, \mathrm{d}v). \tag{12}$$

We now consider the signs of u and v. First, if $u \in [0,1]$ and $v \in [0,1]$, define

$$G_n^{+,+}(t_1, t_2, u, v) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(\boldsymbol{X}_i \in [t_1, t_1 + h_n u] \times [t_2, t_2 + h_n v])I(\Delta_i = 1)}{\bar{H}(\boldsymbol{X}_i-)}.$$

The integrand function in (12) can be deduced as

$$\sup_{t_1,t_2,u\in[0,1],v\in[0,1]}\left|h_n^{-2}\int_{t_1}^{t_1+h_n u}\int_{t_2}^{t_2+h_n v}\frac{1}{\bar{H}(s_1-,s_2-)}(H_{n11} - H_{11})(\mathrm{d}s_1, \mathrm{d}s_2)\right|$$
$$= h_n^{-2}\sup_{t_1,t_2,u\in[0,1],v\in[0,1]}|G_n^{+,+}(t_1,t_2,u,v) - \mathbb{E}\{G_n^{+,+}(t_1,t_2,u,v)\}|. \tag{13}$$

We further define the class of functions

$$\mathscr{G} = \{g_{t_1,t_2,u,v,h_n}(z_1,z_2,z_3)$$
$$= I(z_1 \in [t_1, t_1 + h_n u])I(z_2 \in [t_2, t_2 + h_n v])I(z_3 = 1)\overline{H}^{-1}(z_1-, z_2-) :$$
$$t_1 \in [0, \tau_1], t_2 \in [0, \tau_2], u \in [0,1], v \in [0,1], h_n > 0\}$$

and its subclasses

$$\mathscr{G}_m = \{g_{t_1,t_2,u,v,h}(z_1,z_2,z_3)$$
$$= I(z_1 \in [t_1, t_1 + hu])I(z_2 \in [t_2, t_2 + hv])I(z_3 = 1)\overline{H}^{-1}(z_1-, z_2-) :$$
$$t_1 \in [0, \tau_1], t_2 \in [0, \tau_2], u \in [0,1], v \in [0,1], h_{2^m} \leq h \leq h_{2^{m-1}}\},$$

where $m = 1, 2, \ldots$. It follows form Theorem 3.3 in Giné and Guillou (2001) that $\mathscr{G}$ and, for all $m \geq 1$, $\mathscr{G}_m$ are measurable uniformly bounded VC classes of functions. We now verify conditions in Corollary 2.2 in Giné and Guillou (2002). Due to the boundedness of $\bar{H}^{-1}$ on $[\mathbf{0}, \boldsymbol{\tau}^*]$ in assumption (i), there exists some constant c_6 such that

$$\sup_{g\in\mathscr{G}_m}\sup_{z_1,z_2,z_3}|g(z_1,z_2,z_3)| \leq \sup_{\boldsymbol{s}\in[\mathbf{0},\boldsymbol{\tau}^*]}\bar{H}^{-1}(\boldsymbol{s}) \leq c_6.$$

Furthermore, some calculations, combined with (10), lead to

$$\sup_{t_1\in[0,\tau_1],t_2\in[0,\tau_2],u\in[0,1],v\in[0,1],h_{2m}\le h\le h_{2m-1}} \mathbb{E}\left\{\frac{I(\boldsymbol{X}\in[t_1,t_1+hu]\times[t_2,t_2+hv])I(\Delta=1)}{\bar{H}(\boldsymbol{X}-)}\right\}^2$$
$$\le \sup_{t_1\in[0,\tau_1],t_2\in[0,\tau_2],u\in[0,1],v\in[0,1],h_{2m}\le h\le h_{2m-1}} \int_{t_1}^{t_1+hu}\int_{t_2}^{t_2+hv} H''_{11}(s_1,s_2)\,\mathrm{d}s_1\mathrm{d}s_2$$
$$\times \sup_{\boldsymbol{s}\in[\boldsymbol{0},\boldsymbol{\tau}^*]} \bar{H}^{-2}(\boldsymbol{s}) \le c_6^2c_4 \sup_{h_{2m}\le h\le h_{2m-1}} h^2$$
$$\le c_6^2c_4h_{2m-1}^2.$$

Thus, the parameters U_m and σ_m^2 in Corollary 2.2 in Giné and Guillou (2002) can be taken to be

$$U_m = c_6 \quad \text{and} \quad \sigma_m^2 = c_6^2c_4h_{2m-1}^2.$$

Assumption (iii) implies $h_{2^{m-1}} \to 0$ and $2^{m-1}h_{2^{m-1}}^2/|\log h_{2^{m-1}}| \to \infty$ as $m\to\infty$. Therefore, for sufficiently large m, it holds

$$\sigma_m < U_m/2 \quad \text{and} \quad \sqrt{2^m}\sigma_m \ge U_m\sqrt{\log\frac{U_m}{\sigma_m}}. \tag{14}$$

Similarly, for sufficiently large m, there exists some c_7 such that

$$\sigma_m\sqrt{2^m}\sqrt{\log\frac{U_m}{\sigma_m}} \le \sqrt{c_72^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}. \tag{15}$$

Thus, conditions in Corollary 2.2 in Giné and Guillou (2002) have been verified from (14) and (15), which immediately implies that

$$\mathbb{P}\left(\sup_{g\in\mathscr{G}_m}\left|\sum_{i=1}^{2^m}[g(X_{i1},X_{i2},\Delta_i)-\mathbb{E}\{g(X_{11},X_{12},\Delta_1)\}]\right| > c_8\sqrt{c_72^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}\right)$$
$$\le c_9\exp\left\{-c_{10}\log\frac{U_m}{\sigma_m}\right\} \tag{16}$$

for some positive constants c_8, c_9 and c_{10}, where$\boldsymbol{X}_i=(X_{i1},X_{i2})^{\mathrm{T}}$. Assumption (iii) implies that $|\log h_{2^{m-1}}|/\log\log(2^{m-1})\to\infty$, from which we have

$$c_{10}\log\frac{U_m}{\sigma_m} \ge 2\log(m-1)$$

for sufficiently large m. Then (16) can be further rewritten as

$$\mathbb{P}\left(\sup_{g\in\mathscr{G}_m}\left|\sum_{i=1}^{2^m}[g(X_{i1},X_{i2},\Delta_i)-\mathbb{E}\{g(X_{11},X_{12},\Delta_1)\}]\right| > c_8\sqrt{c_72^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}\right)$$

$$\leq c_9 \exp\{-2\log(m-1)\}$$
$$\leq \frac{c_9}{(m-1)^2}. \tag{17}$$

For $2^{m-1} \leq n \leq 2^m$, we have $h_{2^m} \leq h_n \leq h_{2^{m-1}}$ and $nh_n^2 \geq 2^{m-1}h_{2^{m-1}}^2$ under assumption (iii). As a result, $|\log h_n| \geq |\log h_{2^{m-1}}|$, which immediately implies that

$$\frac{1}{\sqrt{nh_n^2|\log h_n|}} \leq \frac{1}{\sqrt{2^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}}. \tag{18}$$

Following (17), (18) and the Montgomery–Smith maximal inequality (Giné and Guillou 2002), we have

$$\begin{aligned}
&\mathbb{P}\Bigg(\max_{2^{m-1}\leq n\leq 2^m} \sup_{t_1,t_2,u\in[0,1],v\in[0,1]} \sqrt{\frac{n}{h_n^2|\log h_n|}}|G_n^{+,+}(t_1,t_2,u,v) \\
&\quad -\mathbb{E}\{G_n^{+,+}(t_1,t_2,u,v)\}| > 30c_8\sqrt{c_7}\Bigg) \\
&= \mathbb{P}\Bigg(\max_{2^{m-1}\leq n\leq 2^m} \sup_{t_1,t_2,u\in[0,1],v\in[0,1]} \frac{1}{\sqrt{nh_n^2|\log h_n|}} \\
&\quad \times \Bigg|\sum_{i=1}^{n}\Bigg\{\frac{I(\boldsymbol{X}_i\in[t_1,t_1+h_nu]\times[t_2,t_2+h_nv])I(\Delta_i=1)}{\bar{H}(\boldsymbol{X}_i-)} \\
&\quad - \int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{1}{\bar{H}(s_1-,s_2-)}H_{11}(\mathrm{d}s_1,\mathrm{d}s_2)\Bigg\}\Bigg| > 30c_8\sqrt{c_7}\Bigg) \\
&\leq \mathbb{P}\Bigg(\max_{2^{m-1}\leq n\leq 2^m} \sup_{t_1,t_2,u\in[0,1],v\in[0,1],h_{2^m}\leq h\leq h_{2^{m-1}}} \frac{1}{\sqrt{2^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}} \\
&\quad \times \Bigg|\sum_{i=1}^{n}\Bigg\{\frac{I(\boldsymbol{X}_i\in[t_1,t_1+hu]\times[t_2,t_2+hv])I(\Delta_i=1)}{\bar{H}(\boldsymbol{X}_i-)} \\
&\quad - \int_{t_1}^{t_1+hu}\int_{t_2}^{t_2+hv}\frac{1}{\bar{H}(s_1-,s_2-)}H_{11}(\mathrm{d}s_1,\mathrm{d}s_2)\Bigg\}\Bigg| > 30c_8\sqrt{c_7}\Bigg) \\
&\leq 9\mathbb{P}\Bigg(\sup_{t_1,t_2,u\in[0,1],v\in[0,1],h_{2^m}\leq h\leq h_{2^{m-1}}} \frac{1}{\sqrt{2^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}} \\
&\quad \times \Bigg|\sum_{i=1}^{2^m}\Bigg\{\frac{I(\boldsymbol{X}_i\in[t_1,t_1+hu]\times[t_2,t_2+hv])I(\Delta_i=1)}{\bar{H}(\boldsymbol{X}_i-)} \\
&\quad - \int_{t_1}^{t_1+hu}\int_{t_2}^{t_2+hv}\frac{1}{\bar{H}(s_1-,s_2-)}H_{11}(\mathrm{d}s_1,\mathrm{d}s_2)\Bigg\}\Bigg| > c_8\sqrt{c_7}\Bigg)
\end{aligned}$$

$$= 9\mathbb{P}\left(\sup_{g\in\mathscr{G}_m}\left|\sum_{i=1}^{2^m}[g(X_{i1},X_{i2},\Delta_i)-\mathbb{E}\{g(X_{11},X_{12},\Delta_1)\}]\right|\right.$$
$$\left.> c_8\sqrt{c_7 2^{m-1}h_{2^{m-1}}^2|\log h_{2^{m-1}}|}\right)$$
$$\le \frac{9c_9}{(m-1)^2}.$$

Consequently, it follows from the Borel–Cantelli and the zero-one law that with probability one it holds

$$\sqrt{\frac{n}{h_n^2|\log h_n|}}\sup_{t_1,t_2,u\in[0,1],v\in[0,1]}|G_n^{+,+}(t_1,t_2,u,v)-\mathbb{E}\{G_n^{+,+}(t_1,t_2,u,v)\}|\le 30c_8\sqrt{c_7}.$$

As a result,

$$\sup_{t_1,t_2,u\in[0,1],v\in[0,1]}|G_n^{+,+}(t_1,t_2,u,v)-\mathbb{E}\{G_n^{+,+}(t_1,t_2,u,v)\}| = O\left(\sqrt{\frac{h_n^2|\log h_n|}{n}}\right). \quad (19)$$

Then (13) can be rewritten as

$$\sup_{t_1,t_2,u\in[0,1],v\in[0,1]}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{1}{\bar{H}(s_1-,s_2-)}(H_{n11}-H_{11})(\mathrm{d}s_1,\mathrm{d}s_2)\right|$$
$$= O\left(\sqrt{\frac{|\log h_n|}{nh_n^2}}\right).$$

Furthermore, $G_n^{+,-}$, $G_n^{-,+}$, and $G_n^{-,-}$ can be defined similarly. Mimicking the proof of (19), we can also establish the same uniform convergence rate to their expectations. Hence, we conclude that

$$\sup_{t_1,t_2,u\in[-1,1],v\in[-1,1]}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{1}{\bar{H}(s_1-,s_2-)}(H_{n11}-H_{11})(\mathrm{d}s_1,\mathrm{d}s_2)\right|$$
$$= O\left(\sqrt{\frac{|\log h_n|}{nh_n^2}}\right).$$

Based on (12), we have

$$\sup_{\boldsymbol{t}}|\eta_{n2}(\boldsymbol{t})|$$
$$\le \sup_{t_1,t_2,u\in[-1,1],v\in[-1,1]}\left|h_n^{-2}\int_{t_1}^{t_1+h_nu}\int_{t_2}^{t_2+h_nv}\frac{1}{\bar{H}(s_1-,s_2-)}(H_{n11}-H_{11}(\mathrm{d}s_1,\mathrm{d}s_2)\right|\|K\|_{\mathrm{TV}}$$
$$= O\left(\sqrt{\frac{|\log h_n|}{nh_n^2}}\right). \quad (20)$$

Coupled with (11) and (20), (3) can be deduced as

$$\begin{aligned}
\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} &= \sup_{t\in[0,\tau]} |\widehat{\lambda}_n(t) - \lambda_n(t)| \\
&\le \sup_{t\in[0,\tau]} |\eta_{n1}(t)| + \sup_{t\in[0,\tau]} |\eta_{n2}(t)| \\
&\le O\left(\frac{\log\log n}{nh_n^2} + \sqrt{\frac{\log\log n}{n}} + \sqrt{\frac{|\log h_n|}{nh_n^2}}\right) \\
&= O\left(\sqrt{\frac{|\log h_n|}{nh_n^2}}\right)
\end{aligned}$$

under assumption (iii). For scenarios $k = 1$ and 2 and $d \ge 2$, using the similar arguments, we can conclude that

$$\|\widehat{\lambda}_n - \lambda_n\|^*_{k,\max} = O\left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2k}}}\right),$$

which completes the proof of Theorem 3.1. ■

Proof of Theorem 3.2: Assumption (iv) implies that there exist some positive constants c_{11} and c_{12} such that if $|\ell_1 - \ell_2| \le c_{11}$ then

$$\mathsf{Haus}(\{t \in [0, \tau_0] : \lambda_n(t) = \ell_1\}, \{t \in [0, \tau_0] : \lambda_n(t) = \ell_2\}) \le c_{12}|\ell_1 - \ell_2|; \qquad (21)$$

see remarks of Theorem 2 in Cuevas et al. (2006). We first consider $\sup_{x\in\mathcal{L}_n} \rho(x, \widehat{\mathcal{L}}_n)$. It follows from assumption (iii) and Theorem 3.1 that $\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} \xrightarrow{\text{a.s.}} 0$. Taking $x \in \mathcal{L}_n$ and using (21), for sufficiently large n, there exist u_n and v_n, depending on x and $\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}$, such that

$$\lambda_n(u_n) = \ell + 2\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}, \quad \rho(x, u_n) \le 2c_{12}\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}$$

and

$$\lambda_n(v_n) = \ell - 2\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}, \quad \rho(x, v_n) \le 2c_{12}\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}.$$

Thus, we obtain that

$$\widehat{\lambda}_n(u_n) = \lambda_n(u_n) + \widehat{\lambda}_n(u_n) - \lambda_n(u_n) \ge \ell + 2\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} - \|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} \ge \ell.$$

Using the analogous arguments, we have $\widehat{\lambda}_n(v_n) \le \ell$. Hence, there exists some z_n such that $\widehat{\lambda}_n(z_n) = \ell$ and $\rho(z_n, u_n) \le \rho(u_n, v_n)$, which implies that

$$\rho(x, z_n) \le \rho(x, u_n) + \rho(z_n, u_n) \le \rho(x, u_n) + \rho(x, u_n) + \rho(x, v_n) \le 6c_{12}\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}.$$

Immediately,

$$\sup_{x\in\mathcal{L}_n} \rho(x, \widehat{\mathcal{L}}_n) \le 6c_{12}\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}. \qquad (22)$$

We further consider $\sup_{\widehat{\boldsymbol{x}}\in\widehat{\mathcal{L}}_n} \rho(\widehat{\boldsymbol{x}}, \mathcal{L}_n)$. Taking $\widehat{\boldsymbol{x}} \in \widehat{\mathcal{L}}_n$, we have

$$|\lambda_n(\widehat{\boldsymbol{x}}) - \ell| = |\lambda_n(\widehat{\boldsymbol{x}}) - \widehat{\lambda}_n(\widehat{\boldsymbol{x}})| \leq \|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max},$$

which, combined with (21), implies that

$$\sup_{\widehat{\boldsymbol{x}}\in\widehat{\mathcal{L}}_n} \rho(\widehat{\boldsymbol{x}}, \mathcal{L}_n) \leq c_{12} \sup_{\widehat{\boldsymbol{x}}\in\widehat{\mathcal{L}}_n} |\lambda_n(\widehat{\boldsymbol{x}}) - \ell| \leq c_{12}\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}. \tag{23}$$

Following (22), (23) and Theorem 3.1, we have

$$\begin{aligned}
\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) &= \max\left\{\sup_{\widehat{\boldsymbol{x}}\in\widehat{\mathcal{L}}_n} \rho(\widehat{\boldsymbol{x}}, \mathcal{L}_n), \sup_{\boldsymbol{x}\in\mathcal{L}_n} \rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)\right\} \\
&\leq 6c_{12}\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} \\
&= O\left(\sqrt{\frac{|\log h_n|}{nh_n^d}}\right).
\end{aligned}$$

Consequently, we conclude Theorem 3.2. ■

Proof of Theorem 3.3: By Theorem 3.2 and assumption (iii), we have $\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \xrightarrow{\text{a.s.}} 0$. Thus, for sufficiently large n, the set $\widehat{\mathcal{L}}_n$ is contained in set $[\mathbf{0}, \boldsymbol{\tau}]$. On the other hand, it follows from assumptions (iii) and (v), Lemma 1 in Chen et al. (2017), Theorems 3.1 and 3.2 that there exists a unique $\boldsymbol{x} \in \mathcal{L}_n$ such that $\rho(\widehat{\boldsymbol{x}}, \mathcal{L}_n) = \rho(\widehat{\boldsymbol{x}}, \boldsymbol{x})$ for each $\widehat{\boldsymbol{x}} \in \widehat{\mathcal{L}}_n$, and vice versa. Furthermore, assumption (iv) also holds for $\widehat{\lambda}_n(\boldsymbol{t})$ as $\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \xrightarrow{\text{a.s.}} 0$ and $\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max} \xrightarrow{\text{a.s.}} 0$.

Noting that $\widehat{\lambda}_n(\widehat{\boldsymbol{x}}) = \lambda_n(\boldsymbol{x})$, and expanding $\widehat{\lambda}_n(\boldsymbol{x})$ around $\widehat{\boldsymbol{x}}$, we obtain that

$$\begin{aligned}
|\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x}) - \nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x} - \widehat{\boldsymbol{x}})| &= |\widehat{\lambda}_n(\boldsymbol{x}) - \widehat{\lambda}_n(\widehat{\boldsymbol{x}}) - \nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x} - \widehat{\boldsymbol{x}})| \\
&= O(\rho^2(\boldsymbol{x}, \widehat{\boldsymbol{x}})) + O(\|\widehat{\lambda}_n - \lambda_n\|^*_{2,\max})\rho^2(\boldsymbol{x}, \widehat{\boldsymbol{x}})
\end{aligned}$$

under assumptions (ii) and (iv). It follows from the Lagrange multiplier method and assumption (v) that

$$|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x} - \widehat{\boldsymbol{x}})| = \|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})\|_2\rho(\boldsymbol{x}, \widehat{\boldsymbol{x}}).$$

As a consequence,

$$\begin{aligned}
&\big|\,|\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x})| - \|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})\|_2\rho(\boldsymbol{x}, \widehat{\boldsymbol{x}})\big| \\
&\quad\leq |\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x}) - \nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})^{\mathrm{T}}(\boldsymbol{x} - \widehat{\boldsymbol{x}})| \\
&\quad= O(\rho^2(\boldsymbol{x}, \widehat{\boldsymbol{x}})) + O(\|\widehat{\lambda}_n - \lambda_n\|^*_{2,\max})\rho^2(\boldsymbol{x}, \widehat{\boldsymbol{x}}).
\end{aligned} \tag{24}$$

On the other hand, the Taylor expansion of $\nabla\lambda_n(\widehat{\boldsymbol{x}})$ around $\boldsymbol{x}$ results in

$$\nabla\lambda_n(\widehat{\boldsymbol{x}}) - \nabla\lambda_n(\boldsymbol{x}) = \nabla^2\lambda_n(\boldsymbol{x})(\widehat{\boldsymbol{x}} - \boldsymbol{x}) + o(\rho(\widehat{\boldsymbol{x}}, \boldsymbol{x})).$$

Using assumptions (ii), (iv) and Theorem 3.2, we obtain that

$$\|\nabla\lambda_n(\widehat{\boldsymbol{x}}) - \nabla\lambda_n(\boldsymbol{x})\|_2 \le O(\rho(\widehat{\boldsymbol{x}}, \boldsymbol{x})) \le O(\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n)) = O(\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}).$$

Therefore,

$$\begin{aligned}
|\,\|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})\|_2 - \|\nabla\lambda_n(\boldsymbol{x})\|_2| &\le \|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}}) - \nabla\lambda_n(\boldsymbol{x})\|_2 \\
&\le \|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}}) - \nabla\lambda_n(\widehat{\boldsymbol{x}})\|_2 + \|\nabla\lambda_n(\widehat{\boldsymbol{x}}) - \nabla\lambda_n(\boldsymbol{x})\|_2 \\
&\le O(\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max}) + O(\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}) \\
&= O(\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max}).
\end{aligned}$$

Thus, based on Theorems 3.1, 3.2 and assumption (iii), from (24) we have

$$\begin{aligned}
&|\,|\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x})| - \|\nabla\lambda_n(\boldsymbol{x})\|_2\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)| \\
&\quad \le |\,|\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x})| - \|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})\|_2\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)| + |\|\nabla\widehat{\lambda}_n(\widehat{\boldsymbol{x}})\|_2 - \|\nabla\lambda_n(\boldsymbol{x})\|_2|\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n) \\
&\quad = \rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)\{O(\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}) + O(\|\widehat{\lambda}_n - \lambda_n\|^*_{2,\max})O(\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max}) \\
&\qquad + O(\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max})\} \\
&\quad = O(\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max})\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n).
\end{aligned}$$

Since

$$\frac{\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x})}{\|\nabla\lambda_n(\boldsymbol{x})\|_2} = \frac{1}{\sqrt{nh_n^d}}\mathbb{G}_n(f_{\boldsymbol{x}})$$

and $\|\nabla\lambda_n(\boldsymbol{x})\|_2$ is bounded away from zero as stated in assumption (iv), we obtain that

$$\left|\frac{|\mathbb{G}_n(f_{\boldsymbol{x}})| - \sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)}{\sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)}\right| = O(\|\widehat{\lambda}_n - \lambda_n\|^*_{1,\max}) = O\left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2}}}\right)$$

holds uniformly over $\boldsymbol{x} \in \mathcal{L}_n$, which proves Theorem 3.3. ■

Before proving Theorem 3.4, we establish the following lemma, which is a strengthened version of Theorem 3.1.

Lemma 6.1: *Suppose that assumptions* (i)–(iv) *hold. Then for* $k = 0, 1, 2$, *we have*

$$\left|\|\widehat{\lambda}_n - \lambda_n\|^*_{k,\max} - \|g_{n11} - \mathbb{E}(g_{n11})\|^*_{k,\max}\right| = O\left(\frac{\log\log n}{nh_n^{d+k}} + \sqrt{\frac{\log\log n}{nh_n^{2k}}}\right)$$

and

$$\|\widehat{\lambda}_n - \lambda_n\|^*_{k,\max} = O(\|g_{n11} - \mathbb{E}(g_{n11})\|^*_{k,\max}),$$

where

$$g_{n11}(\boldsymbol{t}) = \frac{1}{n}\sum_{i=1}^{n} K_n(\boldsymbol{t} - \boldsymbol{X}_i) I(\Delta_i = 1)\bar{H}^{-1}(\boldsymbol{X}_i-).$$

Proof: We first consider case $k = 0$ and $d = 2$. Under assumption (ii) and using integration by parts, $\eta_{n2}(\boldsymbol{t})$ in (3) can be written as

$$\eta_{n2}(\boldsymbol{t}) = \int_{\mathcal{B}_n(\boldsymbol{t})} \frac{K_n(\boldsymbol{t} - \boldsymbol{s})}{\bar{H}(\boldsymbol{s}-)}(H_{n11} - H_{11})(\mathrm{d}\boldsymbol{s}) = g_{n11}(\boldsymbol{t}) - \mathbb{E}\{g_{n11}(\boldsymbol{t})\}.$$

Following (3) and (11), we obtain that

$$\begin{aligned}
&\left|\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} - \|g_{n11} - \mathbb{E}(g_{n11})\|^*_{0,\max}\right| \\
&= \left|\sup_{\boldsymbol{t}\in[\boldsymbol{0},\boldsymbol{\tau}]} |\widehat{\lambda}_n(\boldsymbol{t}) - \lambda_n(\boldsymbol{t})| - \sup_{\boldsymbol{t}\in[\boldsymbol{0},\boldsymbol{\tau}]} |g_{n11}(\boldsymbol{t}) - \mathbb{E}\{g_{n11}(\boldsymbol{t})\}|\right| \\
&\le \sup_{\boldsymbol{t}\in[\boldsymbol{0},\boldsymbol{\tau}]} |\widehat{\lambda}_n(\boldsymbol{t}) - \lambda_n(\boldsymbol{t}) - \eta_{n2}(\boldsymbol{t})| \\
&= \sup_{\boldsymbol{t}\in[\boldsymbol{0},\boldsymbol{\tau}]} |\eta_{n1}(\boldsymbol{t})| \\
&= O\left(\frac{\log\log n}{nh_n^2} + \sqrt{\frac{\log\log n}{n}}\right).
\end{aligned}$$

Using the analogous arguments to Theorem 3.3 in Giné and Guillou (2002), there exists some constant c_{13} such that $\|g_{n11} - \mathbb{E}(g_{n11})\|^*_{0,\max} \ge c_{13}\sqrt{|\log h_n|/(nh_n^d)}$. Thus, assumption (iii) implies that

$$\begin{aligned}
\|\widehat{\lambda}_n - \lambda_n\|^*_{0,\max} &\le \|g_{n11} - \mathbb{E}(g_{n11})\|^*_{0,\max} + O\left(\frac{\log\log n}{nh_n^d} + \sqrt{\frac{\log\log n}{n}}\right) \\
&= O(\|g_{n11} - \mathbb{E}(g_{n11})\|^*_{0,\max}).
\end{aligned}$$

For scenarios $k = 1, 2$ and $d \ge 2$, utilising the similar arguments, we conclude Lemma 6.1. ■

Proof of Theorem 3.4: Based on Theorems 3.2 and 3.3 and some basic derivations, we have

$$\left|\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) - \sup_{f\in\mathscr{F}} |\mathbb{G}_n(f)|\right|$$

$$
\begin{aligned}
&= \left| \sqrt{nh_n^d} \sup_{\boldsymbol{x}\in\mathcal{L}_n} \rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n) - \sup_{f\in\mathscr{F}} |\mathbb{G}_n(f)| \right| \\
&\le \sup_{\boldsymbol{x}\in\mathcal{L}_n} \sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n) \frac{\sup_{\boldsymbol{x}\in\mathcal{L}_n} \left| |\mathbb{G}_n(f_{\boldsymbol{x}})| - \sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n) \right|}{\sup_{\boldsymbol{x}\in\mathcal{L}_n} \sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)} \\
&\le \sup_{\boldsymbol{x}\in\mathcal{L}_n} \sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n) \sup_{\boldsymbol{x}\in\mathcal{L}_n} \left| \frac{|\mathbb{G}_n(f_{\boldsymbol{x}})| - \sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)}{\sqrt{nh_n^d}\rho(\boldsymbol{x}, \widehat{\mathcal{L}}_n)} \right| \\
&\le c_{14}\sqrt{|\log h_n|}\|\widehat{\lambda}_n - \lambda_n\|_{1,\max}^*
\end{aligned}
$$

for some positive constant c_{14}. By Corollary 2.2 in Giné and Guillou (2002), there exist some constants c_{15}, c_{16}, c_{17} and c_{18} such that for $c_{15}\sqrt{|\log h_n|/(nh_n^{d+2})} \le t_1 \le c_{16}$,

$$
\mathbb{P}(\|g_{n11} - \mathbb{E}(g_{n11})\|_{1,\max}^* > t_1) \le c_{17} \exp(-c_{18} nh_n^{d+2} t_1^2). \tag{25}
$$

Following Lemma 6.1, we have

$$
\begin{aligned}
&\mathbb{P}\left(\left| \sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) - \sup_{f\in\mathscr{F}} |\mathbb{G}_n(f)| \right| > t_2 \right) \\
&\quad\le \mathbb{P}(c_{14}\sqrt{|\log h_n|}\|\widehat{\lambda}_n - \lambda_n\|_{1,\max}^* > t_2) \\
&\quad\le \mathbb{P}(c_{19}c_{14}\sqrt{|\log h_n|}\|g_{n11} - \mathbb{E}(g_{n11})\|_{1,\max}^* > t_2) \\
&\quad\le c_{17} \exp\left(-\frac{c_{18}}{c_{19}^2 c_{14}^2} \frac{nh_n^{d+2}}{|\log h_n|} t_2^2 \right)
\end{aligned} \tag{26}
$$

for some positive constant c_{19} and $c_{15}c_{19}c_{14}\sqrt{|\log h_n|^2/(nh_n^{d+2})} \le t_2 \le c_{16}c_{19}c_{14}\sqrt{|\log h_n|}$.

On the other hand, we define the class of functions

$$
\mathscr{F}^\dagger = \left\{ f_{\boldsymbol{x}}(\boldsymbol{z}_1, z_2) = \frac{K((\boldsymbol{x} - \boldsymbol{z}_1)/h_n) I(z_2 = 1)}{\sqrt{h_n^d}\|\nabla\lambda_n(\boldsymbol{x})\|_2 \bar{H}(\boldsymbol{z}_1 -)} : \boldsymbol{x} \in \mathcal{L}_n \right\},
$$

on which we further define the process

$$
\mathbb{G}_n^\dagger(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [f(\boldsymbol{X}_i, \Delta_i) - \mathbb{E}\{f(\boldsymbol{X}_i, \Delta_i)\}]
$$

for any $f \in \mathscr{F}^\dagger$. We denote the Gaussian process $\mathbb{B}^\dagger$ indexed by $\mathscr{F}^\dagger$ such that for any $f_1 \in \mathscr{F}^\dagger$ and $f_2 \in \mathscr{F}^\dagger$,

$$
\mathbb{B}^\dagger(f_1) \stackrel{D}{=} N(0, \mathbb{E}\{f_1^2(\boldsymbol{X}, \Delta)\}) \quad \text{and} \quad \mathrm{Cov}(\mathbb{B}^\dagger(f_1), \mathbb{B}^\dagger(f_2)) = \mathbb{E}\{f_1(\boldsymbol{X}, \Delta) f_2(\boldsymbol{X}, \Delta)\}.
$$

Following Corollary 2.2 in Chernozhukov, Chetverikov, and Kato (2014), there exist some positive constants c_{20} and c_{21} and some random variable $\boldsymbol{B} \stackrel{D}{=} \sup_{f\in\mathscr{F}^\dagger} |\mathbb{B}^\dagger(f)|$ such that for all $\gamma \in (0,1)$ and sufficiently large n,

$$\mathbb{P}\left(\left|\sup_{f\in\mathscr{F}^\dagger}|\mathbb{G}_n^\dagger(f)| - \boldsymbol{B}\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right) \le c_{21}\gamma.$$

Obviously,

$$\begin{aligned}
&\mathbb{P}\left(\left|\sup_{f\in\mathscr{F}}|\mathbb{G}_n(f)| - \boldsymbol{B}\right| > c_{20}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right)\\
&\quad\le \mathbb{P}\left(\left|\sup_{f\in\mathscr{F}^\dagger}|\mathbb{G}_n^\dagger(f)| - \boldsymbol{B}\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right)\\
&\qquad+ \mathbb{P}\left(\left|\sup_{f\in\mathscr{F}}|\mathbb{G}_n(f)| - \sup_{f\in\mathscr{F}^\dagger}|\mathbb{G}_n^\dagger(f)|\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right)\\
&\quad\le c_{21}\gamma + \mathbb{P}\left(\left|\sup_{f\in\mathscr{F}}|\mathbb{G}_n(f)| - \sup_{f\in\mathscr{F}^\dagger}|\mathbb{G}_n^\dagger(f)|\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right).
\end{aligned} \tag{27}$$

By (26) and (27), we have

$$\begin{aligned}
&\mathbb{P}\left(|\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n,\mathcal{L}_n) - \boldsymbol{B}| > t_2 + c_{20}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right)\\
&\quad\le c_{17}\exp\left(-\frac{c_{18}}{c_{19}^2c_{14}^2}\frac{nh_n^{d+2}}{|\log h_n|}t_2^2\right) + c_{21}\gamma\\
&\qquad+ \mathbb{P}\left(\left|\sup_{f\in\mathscr{F}}|\mathbb{G}_n(f)| - \sup_{f\in\mathscr{F}^\dagger}|\mathbb{G}_n^\dagger(f)|\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right).
\end{aligned} \tag{28}$$

On the other hand, viewed as index by $\boldsymbol{x}$ over $\mathcal{L}_n$, the distribution of $\mathbb{B}(f)$ is the same as that of $\mathbb{B}^\dagger(f)$. Consequently, employing Lemma 10 in Chen et al. (2017) and (28), we have

$$\begin{aligned}
&\sup_t\left|\mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n,\mathcal{L}_n) \le t) - \mathbb{P}\left(\sup_{f\in\mathscr{F}}|\mathbb{B}(f)| \le t\right)\right|\\
&\quad\le c_{22}\left\{c_{20}\frac{\log^{7/6} n}{\gamma^{1/3}(nh_n^d)^{1/6}} + t_2\sqrt{\log n}\right\} + c_{21}\gamma + c_{17}\exp\left(-\frac{c_{18}}{c_{19}^2c_{14}^2}\frac{nh_n^{d+2}}{|\log h_n|}t_2^2\right)\\
&\qquad+ \mathbb{P}\left(\left|\sup_{f\in\mathscr{F}}|\mathbb{G}_n(f)| - \sup_{f\in\mathscr{F}^\dagger}|\mathbb{G}_n^\dagger(f)|\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right)
\end{aligned} \tag{29}$$

for some positive constant c_{22}. Observing that

$$\mathbb{G}_n(f_{\boldsymbol{x}}) = \sqrt{nh_n^d}\frac{\widehat{\lambda}_n(\boldsymbol{x}) - \lambda_n(\boldsymbol{x})}{\|\nabla\lambda_n(\boldsymbol{x})\|_2} \quad \text{for } f_{\boldsymbol{x}} \in \mathscr{F} \tag{30}$$

and

$$\mathbb{G}_n^{\dagger}(f_{\boldsymbol{x}}) = \sqrt{nh_n^d}\frac{g_{n11}(\boldsymbol{x}) - \mathbb{E}\{g_{n11}(\boldsymbol{x})\}}{\|\nabla\lambda_n(\boldsymbol{x})\|_2} \quad \text{for } f_{\boldsymbol{x}} \in \mathscr{F}^{\dagger}, \tag{31}$$

and using Lemma 6.1 and the boundedness of $\|\nabla\lambda_n(\boldsymbol{x})\|_2^{-1}$ over $\boldsymbol{x} \in \mathcal{L}_n$, which is stated in assumption (iv), we obtain that

$$\left|\sup_{f\in\mathscr{F}} |\mathbb{G}_n(f)| - \sup_{f\in\mathscr{F}^{\dagger}} |\mathbb{G}_n^{\dagger}(f)|\right| = O\left(\frac{\log\log n}{\sqrt{nh_n^d}} + \sqrt{h_n^d \log\log n}\right).$$

Under assumption (iii) and as $nh_n^{5d}(\log\log n)^4/\log^3 n \to 0$, there exists some $c_{23} \geq c_{15}^2 c_{18}$ large enough such that $nh_n^{d+2+2c_{23}}/(|\log h_n|^2 \log n)$ converges to zero. By setting $\gamma = (\frac{\log^7 n}{nh_n^d})^{1/8}$ and $t_2 = (\frac{c_{23}c_{19}^2 c_{14}^2 |\log h_n|^2}{c_{18}nh_n^{d+2}})^{1/2}$, it is straightforward to verify that $\frac{\log\log n}{\sqrt{nh_n^d}}$ and $\sqrt{h_n^d \log\log n}$ converge to zero faster than $\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}$. Therefore, we have

$$\mathbb{P}\left(\left|\sup_{f\in\mathscr{F}} |\mathbb{G}_n(f)| - \sup_{f\in\mathscr{F}^{\dagger}} |\mathbb{G}_n^{\dagger}(f)|\right| > \frac{c_{20}}{2}\frac{\log^{2/3} n}{\gamma^{1/3}(nh_n^d)^{1/6}}\right) = 0. \tag{32}$$

Based on the chosen γ and t_2, (29) can be refined as

$$\begin{aligned} &\sup_t \left|\mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \leq t) - \mathbb{P}\left(\sup_{f\in\mathscr{F}} |\mathbb{B}(f)| \leq t\right)\right| \\ &= O\left(\left(\frac{\log^7 n}{nh_n^d}\right)^{1/8} + \left(\frac{\log^3 n}{nh_n^{d+2}}\right)^{1/2}\right) \end{aligned}$$

by using Remark 3.1. We therefore complete the proof of Theorem 3.4. ■

Proof of Theorem 3.5: Denote $\mathscr{K}_n = \{(\mathcal{O}_1, \dots, \mathcal{O}_n) : \|\widehat{\lambda}_n - \lambda_n\|_{1,\max}^* \leq \varepsilon_0\}$ for some small ε_0. As a result, whenever dataset lie in $\mathscr{K}_n$, assumption (iv) holds for $\widehat{\lambda}_n$ by replacing $\mathcal{L}_n$ with $\widehat{\mathcal{L}}_n$. In addition, by Lemma 6.1 and (25), we have $\mathbb{P}(\mathscr{K}_n) \geq 1 - c_{24}\exp(-c_{25}nh_n^{d+2})$ for some constants c_{24} and c_{25}. Thus, we can assume that the original data $\mathcal{O}_1, \dots, \mathcal{O}_n$ lies in $\mathscr{K}_n$. We define the bootstrap empirical process

$$\mathbb{G}_n^*(f) = \int_{[\mathbf{0},\boldsymbol{\tau}_0]} f(\boldsymbol{s})\, \mathrm{d}[n^{1/2}\{\widehat{\Lambda}^*(\boldsymbol{s}) - \widehat{\Lambda}(\boldsymbol{s})\}],$$

where $\widehat{\Lambda}^*(\boldsymbol{s})$ is the bootstrapped estimator. Let $\widehat{\mathcal{L}}_n^*$ be the bootstrapped estimator of hazard level set, then we have

$$\begin{aligned} &\sup_t |\mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n^*, \widehat{\mathcal{L}}_n) \leq t|\mathcal{O}_1, \dots, \mathcal{O}_n) - \mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \leq t)| \\ &\leq \sup_t \left|\mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n^*, \widehat{\mathcal{L}}_n) \leq t|\mathcal{O}_1, \dots, \mathcal{O}_n) - \mathbb{P}\left(\sup_{f\in\mathscr{F}_n^*} |\mathbb{B}_n^*(f)| \leq t|\mathcal{O}_1, \dots, \mathcal{O}_n\right)\right| \end{aligned}$$

$$+ \sup_t \left| \mathbb{P}\left(\sup_{f \in \mathscr{F}} |\mathbb{B}(f)| \leq t \right) - \mathbb{P}\left(\sup_{f \in \mathscr{F}_n^*} |\mathbb{B}_n^*(f)| \leq t | \mathcal{O}_1, \ldots, \mathcal{O}_n \right) \right|$$

$$+ \sup_t \left| \mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \leq t) - \mathbb{P}\left(\sup_{f \in \mathscr{F}} |\mathbb{B}(f)| \leq t \right) \right|$$

$$= \zeta_{n1} + \zeta_{n2} + \zeta_{n3}, \tag{33}$$

where ζ_{n1}, ζ_{n2} and ζ_{n3} are self-explained from the expression,

$$\mathscr{F}_n^* = \left\{ f_{\boldsymbol{x}}(\cdot) = \frac{1}{\sqrt{h_n^d}\|\nabla\widehat{\lambda}_n(\boldsymbol{x})\|_2} K\left(\frac{\boldsymbol{x} - \cdot}{h_n}\right) : \boldsymbol{x} \in \widehat{\mathcal{L}}_n \right\},$$

and $\mathbb{B}_n^*$ is a Gaussian process on $\mathscr{F}_n^*$ such that for any $f_1 \in \mathscr{F}_n^*$ and $f_2 \in \mathscr{F}_n^*$,

$$\mathbb{E}(\mathbb{B}_n^*(f_1)|\mathcal{O}_1, \ldots, \mathcal{O}_n) = 0$$

and

$$\mathrm{Cov}(\mathbb{B}_n^*(f_1), \mathbb{B}_n^*(f_2)|\mathcal{O}_1, \ldots, \mathcal{O}_n) = \frac{1}{n}\sum_{i=1}^n f_1(\boldsymbol{X}_i)f_2(\boldsymbol{X}_i)I(\Delta_i = 1)\bar{H}^{-2}(\boldsymbol{X}_i-).$$

Theorem 3.4 implies that

$$\zeta_{n3} = O\left(\left(\frac{\log^7 n}{nh_n^d} \right)^{1/8} + \left(\frac{\log^3 n}{nh_n^{d+2}} \right)^{1/2} \right). \tag{34}$$

If we suggest to estimate $\widehat{\lambda}_n$ by $\widehat{\lambda}_n^*$, mimicking analogous procedures of estimating λ_n by $\widehat{\lambda}_n$ and applying similar arguments for $\mathsf{Haus}(\widehat{\mathcal{L}}_n^*, \widehat{\mathcal{L}}_n)$, we have

$$\zeta_{n1} = O\left(\left(\frac{\log^7 n}{nh_n^d} \right)^{1/8} + \left(\frac{\log^3 n}{nh_n^{d+2}} \right)^{1/2} \right). \tag{35}$$

Now we consider the second term ζ_{n2}. By assumptions (ii) and (iv), there exists some universal constant c_{26} such that

$$\|\nabla\lambda_n(\boldsymbol{x}_1)\|_2^{-1} K\left(\frac{\boldsymbol{x}_1 - \boldsymbol{x}}{h_n}\right) \leq c_{26}, \quad \|\nabla\widehat{\lambda}_n(\boldsymbol{x}_2)\|_2^{-1} K\left(\frac{\boldsymbol{x}_2 - \boldsymbol{x}}{h_n}\right) \leq c_{26}$$

hold uniformly over $\boldsymbol{x} \in [\boldsymbol{0}, \boldsymbol{\tau}_0]$, $\boldsymbol{x}_1 \in \mathcal{L}_n$ and $\boldsymbol{x}_2 \in \widehat{\mathcal{L}}_n$, respectively. Following (2.10) in Giné and Guillou (2002), the function collection

$$\mathscr{F}_{\text{scale}} = \left\{ f_{\boldsymbol{x}}(\cdot) = \|\nabla\lambda_n(\boldsymbol{x})\|_2^{-1} K\left(\frac{\boldsymbol{x} - \cdot}{h_n}\right) : \boldsymbol{x} \in \mathcal{L}_n \right\}$$

is the uniformly bounded VC class. As a result, there exist some constants c_{27} and c_{28} such that

$$\sup_{\mathbb{Q}} N(\mathscr{F}_{\text{scale}}, L_2(\mathbb{Q}), c_{26}\epsilon) \leq \left(\frac{c_{27}}{\epsilon}\right)^{c_{28}},$$

where $N(\Omega, \varrho, \epsilon)$ denotes the ϵ-covering number of metric space (Ω, ϱ), $\mathbb{Q}$ is the probability measure and the $L_2(\mathbb{Q})$ norm of f is $(\int |f|^2 \, \mathrm{d}\mathbb{Q})^{1/2}$. Thus, we have

$$N_1(\epsilon, n) = \sup_{\mathbb{Q}} N(\mathscr{F}, L_2(\mathbb{Q}), c_{26}\epsilon) \le \left(\frac{c_{27}}{h_n^{d/2}\epsilon} \right)^{c_{28}} \tag{36}$$

by observing the scale transformation from $\mathscr{F}_{\text{scale}}$ to $\mathscr{F}$. Let $\mathbb{L}$ be the Lebesgue measure on $\mathbb{R}^d$. For $\mathbb{Q} = (\tau_{01} \cdots \tau_{0d})^{-1}\mathbb{L}$ and $\epsilon > 0$, we can find a covering set for $\mathscr{F}$ with size $N_1(\epsilon, n)$, which is denoted by $\mathscr{D}_1 = \{f_1^0, \ldots, f_{N_1(\epsilon,n)}^0\}$. Obviously, class $\mathscr{F}$ is indexed by $\boldsymbol{x}$ over $\mathcal{L}_n$. On the contrary, for any $f \in \mathscr{F}$ we denote the corresponding index by $\boldsymbol{\Psi}_1(f) \in \mathcal{L}_n$.

As far as class $\mathscr{F}_n^*$, utilising analogous arguments to (36), there exist some constants c_{29} and c_{30} such that

$$N_2(\epsilon, n) = \sup_{\mathbb{Q}} N(\mathscr{F}_n^*, L_2(\mathbb{Q}), c_{26}\epsilon) \le \left(\frac{c_{29}}{h_n^{d/2}\epsilon} \right)^{c_{30}}.$$

The covering set $\mathscr{D}_2 = \{f_1^{0,*}, \ldots, f_{N_2(\epsilon,n)}^{0,*}\}$ for $\mathscr{F}_n^*$ and the corresponding index $\boldsymbol{\Psi}_2(f^*) \in \widehat{\mathcal{L}}_n$ for any $f^* \in \mathscr{F}_n^*$ are defined similarly.

Let $N(\epsilon, n) = N_1(\epsilon, n) + N_2(\epsilon, n)$ and $\Pi_A(\boldsymbol{x})$ be the projection for $\boldsymbol{x}$ onto a set A. Obviously, for any $f \in \mathscr{F}$, there exists $f^* \in \mathscr{F}_n^*$ such that $\boldsymbol{\Psi}_2(f^*) = \Pi_{\widehat{\mathcal{L}}_n}(\boldsymbol{\Psi}_1(f))$, and vice versa. With slight abuse of notation, we define the mapping by $\Pi_{\mathscr{F}_n^*}(f) = f^*$. Let

$$f_i = \begin{cases} f_i^0, & 1 \le i \le N_1(\epsilon, n), \\ \Pi_{\mathscr{F}}(f_{i-N_1(\epsilon,n)}^{0,*}), & N_1(\epsilon, n) + 1 \le i \le N(\epsilon, n) \end{cases}$$

and

$$f_i^* = \begin{cases} \Pi_{\mathscr{F}_n^*}(f_i^0), & 1 \le i \le N_1(\epsilon, n), \\ f_{i-N_1(\epsilon,n)}^{0,*}, & N_1(\epsilon, n) + 1 \le i \le N(\epsilon, n), \end{cases}$$

which implies that

$$\|\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_2(f_i^*)\|_2 \le \mathsf{Haus}(\mathcal{L}_n, \widehat{\mathcal{L}}_n) \tag{37}$$

for all $1 \le i \le N(\epsilon, n)$. Let $\breve{\mathscr{F}} = \{f_1, \ldots, f_{N(\epsilon,n)}\}$ and $\breve{\mathscr{F}}_n^* = \{f_1^*, \ldots, f_{N(\epsilon,n)}^*\}$, then ζ_{n2} can be deduced as

$$\begin{aligned} \zeta_{n2} \le \sup_t & \left| \mathbb{P}\left(\sup_{f \in \mathscr{F}} |\mathbb{B}(f)| \le t \right) - \mathbb{P}\left(\sup_{f \in \breve{\mathscr{F}}} |\mathbb{B}(f)| \le t \right) \right| \\ & + \sup_t \left| \mathbb{P}\left(\sup_{f \in \breve{\mathscr{F}}} |\mathbb{B}(f)| \le t \right) - \mathbb{P}\left(\sup_{f \in \breve{\mathscr{F}}_n^*} |\mathbb{B}(f)| \le t | \mathcal{O}_1, \ldots, \mathcal{O}_n \right) \right| \\ & + \sup_t \left| \mathbb{P}\left(\sup_{f \in \breve{\mathscr{F}}_n^*} |\mathbb{B}(f)| \le t | \mathcal{O}_1, \ldots, \mathcal{O}_n \right) - \mathbb{P}\left(\sup_{f \in \breve{\mathscr{F}}_n^*} |\mathbb{B}_n^*(f)| \le t | \mathcal{O}_1, \ldots, \mathcal{O}_n \right) \right| \end{aligned}$$

$$+\sup_{t}\left|\mathbb{P}\left(\sup_{f\in\mathscr{F}_n^*}|\mathbb{B}_n^*(f)|\le t|\mathcal{O}_1,\ldots,\mathcal{O}_n\right)-\mathbb{P}\left(\sup_{f\in\breve{\mathscr{F}}_n^*}|\mathbb{B}_n^*(f)|\le t|\mathcal{O}_1,\ldots,\mathcal{O}_n\right)\right|$$
$$=\zeta_{n2}^{(1)}+\zeta_{n2}^{(2)}+\zeta_{n2}^{(3)}+\zeta_{n2}^{(4)}, \tag{38}$$

where definitions of $\zeta_{n2}^{(1)}$, $\zeta_{n2}^{(2)}$, $\zeta_{n2}^{(3)}$ and $\zeta_{n2}^{(4)}$ are also self-explained. We first deal with term $\zeta_{n2}^{(1)}$. For notational simplicity, denote $\gamma_1=(\frac{\log^7 n}{nh_n^d})^{1/8}$. Under assumption (iii) and $nh_n^{5d}(\log\log n)^4/\log^3 n\to 0$, following (27), (30), (31) and (32), there exist some constants c_{31} and c_{32} such that

$$\mathbb{P}\left(\left|\sup_{\boldsymbol{x}\in\mathcal{L}_n}\sqrt{nh_n^d}\|\nabla\lambda_n(\boldsymbol{x})\|_2^{-1}|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|-\sup_{f\in\mathscr{F}}|\mathbb{B}(f)|\right|>\frac{c_{31}}{3}\frac{\log^{2/3}n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\right)$$
$$\le\mathbb{P}\left(\left|\sup_{\boldsymbol{x}\in\mathcal{L}_n}\sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})t|}{\|\nabla\lambda_n(\boldsymbol{x})\|_2}\right.\right.$$
$$\left.\left.-\sup_{\boldsymbol{x}\in\mathcal{L}_n}\sqrt{nh_n^d}\frac{|g_{n11}(\boldsymbol{x})-\mathbb{E}\{g_{n11}(\boldsymbol{x})\}|}{\|\nabla\lambda_n(\boldsymbol{x})\|_2}\right|>\frac{c_{31}}{3}\frac{\log^{2/3}n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\right)+c_{32}\gamma_1$$
$$=c_{32}\gamma_1, \tag{39}$$

where g_{n11} is defined in Lemma 6.1. In addition, (39) also holds for $\boldsymbol{x}\in\breve{\mathcal{L}}_n$ and $f\in\breve{\mathscr{F}}$, where $\breve{\mathcal{L}}_n=\{\Xi_1(f_1),\ldots,\Xi_1(f_{N(\epsilon,n)})\}$. Hence, we have

$$\mathbb{P}\left(\left|\sup_{f\in\mathscr{F}}|\mathbb{B}(f)|-\sup_{f\in\breve{\mathscr{F}}}|\mathbb{B}(f)|\right|>c_{31}\frac{\log^{2/3}n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\right)$$
$$\le\mathbb{P}\left(\left|\sup_{\boldsymbol{x}\in\mathcal{L}_n}\sqrt{nh_n^d}\|\nabla\lambda_n(\boldsymbol{x})\|_2^{-1}|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|-\sup_{f\in\mathscr{F}}|\mathbb{B}(f)|\right|>\frac{c_{31}}{3}\frac{\log^{2/3}n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\right)$$
$$+\mathbb{P}\left(\left|\sup_{\boldsymbol{x}\in\mathcal{L}_n}\sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|}{\|\nabla\lambda_n(\boldsymbol{x})\|_2}-\sup_{\boldsymbol{x}\in\breve{\mathcal{L}}_n}\sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|}{\|\nabla\lambda_n(\boldsymbol{x})\|_2}\right|\right.$$
$$\left.>\frac{c_{31}}{3}\frac{\log^{2/3}n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\right)$$
$$+\mathbb{P}\left(\left|\sup_{\boldsymbol{x}\in\breve{\mathcal{L}}_n}\sqrt{nh_n^d}\|\nabla\lambda_n(\boldsymbol{x})\|_2^{-1}|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|-\sup_{f\in\breve{\mathscr{F}}}|\mathbb{B}(f)|\right|\right.$$
$$\left.>\frac{c_{31}}{3}\frac{\log^{2/3}n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\right)$$
$$\le c_{33}\gamma_1+\mathbb{P}\left(\left|\sup_{\boldsymbol{x}\in\mathcal{L}_n}\sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|}{\|\nabla\lambda_n(\boldsymbol{x})\|_2}-\sup_{\boldsymbol{x}\in\breve{\mathcal{L}}_n}\sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{x})-\lambda_n(\boldsymbol{x})|}{\|\nabla\lambda_n(\boldsymbol{x})\|_2}\right|\right.$$

$$> \frac{c_{31}}{3} \frac{\log^{2/3} n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\Bigg)$$

$$= c_{33}\gamma_1 + \mathbb{P}\Bigg(\Bigg|\sup_{f\in\mathscr{F}} \sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{\Psi}_1(f)) - \lambda_n(\boldsymbol{\Psi}_1(f))|}{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2}$$

$$- \sup_{f\in\breve{\mathscr{F}}} \sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\boldsymbol{\Psi}_1(f)) - \lambda_n(\boldsymbol{\Psi}_1(f))|}{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2}\Bigg| > \frac{c_{31}}{3} \frac{\log^{2/3} n}{\gamma_1^{1/3}(nh_n^d)^{1/6}}\Bigg)$$

$$= c_{33}\gamma_1 + \kappa_n$$

for some constant c_{33}, where κ_n is self-explained from the expression and will be considered later. Noting that $\gamma_1 = (\frac{\log^7 n}{nh_n^d})^{1/8}$ and using Lemma 10 in Chen et al. (2017), we have for some constant c_{34}

$$\zeta_{n2}^{(1)} = \sup_t \left|\mathbb{P}\left(\sup_{f\in\mathscr{F}} |\mathbb{B}(f)| \le t\right) - \mathbb{P}\left(\sup_{f\in\breve{\mathscr{F}}} |\mathbb{B}(f)| \le t\right)\right|$$

$$\le c_{34}c_{31}\frac{\log^{7/6} n}{\gamma_1^{1/3}(nh_n^d)^{1/6}} + c_{33}\gamma_1 + \kappa_n$$

$$\le (c_{34}c_{31} + c_{33})\left(\frac{\log^7 n}{nh_n^d}\right)^{1/8} + \kappa_n.$$

Now we consider the term $\zeta_{n2}^{(2)}$. Define

$$(\mathbb{B}(f_1), \dots, \mathbb{B}(f_{N(\epsilon,n)})) \overset{D}{=} N(0, \boldsymbol{\Sigma}_1), \quad (\mathbb{B}(f_1^*), \dots, \mathbb{B}(f_{N(\epsilon,n)}^*)) \overset{D}{=} N(0, \boldsymbol{\Sigma}_2)$$

and $\delta_{12} = \max_{1\le i,j\le N(\epsilon,n)} |\Sigma_1^{i,j} - \Sigma_2^{i,j}|$, where $\Sigma_k^{i,j}$ is the (i,j)th element of $\boldsymbol{\Sigma}_k$ for $k = 1, 2$ and $i,j = 1, \dots, N(\epsilon, n)$. Employing Gaussian comparison theorem (Chernozhukov, Chetverikov, and Kato 2013), there exists some constant c_{35} such that

$$\zeta_{n2}^{(2)} \le c_{35}\delta_{12}^{1/3}[\max\{1, (N(\epsilon, n)/\delta_{12})\}]^{2/3}.$$

We next discuss the bound of δ_{12}. Observing that

$$\Sigma_1^{i,j} = \int_{[\boldsymbol{0},\boldsymbol{\tau}_0]} h_n^{-d}\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2^{-1}\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_j))\|_2^{-1} \times K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_1(f_j) - \boldsymbol{x}}{h_n}\right) H_{11}''(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$$

and

$$\Sigma_2^{i,j} = \int_{[\boldsymbol{0},\boldsymbol{\tau}_0]} h_n^{-d}\|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2^{-1}\|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_j^*))\|_2^{-1} \times K\left(\frac{\boldsymbol{\Psi}_2(f_i^*) - \boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_2(f_j^*) - \boldsymbol{x}}{h_n}\right) H_{11}''(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x},$$

where H''_{11} is defined in (7), we have

$$
\begin{aligned}
&|\Sigma_1^{i,j} - \Sigma_2^{i,j}| \\
&\leq \|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2^{-1}\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_j))\|_2^{-1} - \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2^{-1}\|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_j^*))\|_2^{-1}| \\
&\quad \times \int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d} K\left(\frac{\boldsymbol{\Psi}_1(f_i)-\boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_1(f_j)-\boldsymbol{x}}{h_n}\right) H''_{11}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&\quad + \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2^{-1}\|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_j^*))\|_2^{-1} h_n^{-d} \\
&\quad \times \int_{[\mathbf{0},\boldsymbol{\tau}_0]} \left| K\left(\frac{\boldsymbol{\Psi}_1(f_i)-\boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_1(f_j)-\boldsymbol{x}}{h_n}\right)\right. \\
&\quad \left. - K\left(\frac{\boldsymbol{\Psi}_2(f_i^*)-\boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_2(f_j^*)-\boldsymbol{x}}{h_n}\right)\right| H''_{11}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.
\end{aligned}
$$

We obtain that H''_{11} is bounded over $[\mathbf{0},\boldsymbol{\tau}_0]$ because (10) can be further strengthened by replacing $[\mathbf{0},\boldsymbol{\tau}^*]$ with $[\mathbf{0},\boldsymbol{\tau}_0]$. Furthermore, assumption (ii) implies that $\int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d} K(\frac{\boldsymbol{\Psi}_1(f_i)-\boldsymbol{x}}{h_n}) K(\frac{\boldsymbol{\Psi}_1(f_j)-\boldsymbol{x}}{h_n}) H''_{11}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ is also bounded. On the other hand, under assumption (ii) and noting that assumption (iv) holds for λ_n and $\widehat{\lambda}_n$, we have

$$
\begin{aligned}
&|\Sigma_1^{i,j} - \Sigma_2^{i,j}| \\
&\leq O(1)\left\{ h_n^{-d} \int_{[\mathbf{0},\boldsymbol{\tau}_0]} \left| K\left(\frac{\boldsymbol{\Psi}_1(f_i)-\boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_1(f_j)-\boldsymbol{x}}{h_n}\right)\right.\right. \\
&\quad \left. - K\left(\frac{\boldsymbol{\Psi}_2(f_i^*)-\boldsymbol{x}}{h_n}\right) K\left(\frac{\boldsymbol{\Psi}_2(f_j^*)-\boldsymbol{x}}{h_n}\right)\right| \mathrm{d}\boldsymbol{x} \\
&\quad \left. + |\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 - \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2| + |\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_j))\|_2 - \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_j^*))\|_2| \right\} \\
&\leq O(1)\left\{ h_n^{-d} \int_{[\mathbf{0},\boldsymbol{\tau}_0]} \left| K\left(\frac{\boldsymbol{\Psi}_1(f_i)-\boldsymbol{x}}{h_n}\right) - K\left(\frac{\boldsymbol{\Psi}_2(f_i^*)-\boldsymbol{x}}{h_n}\right)\right| \mathrm{d}\boldsymbol{x} \right. \\
&\quad + h_n^{-d} \int_{[\mathbf{0},\boldsymbol{\tau}_0]} \left| K\left(\frac{\boldsymbol{\Psi}_1(f_j)-\boldsymbol{x}}{h_n}\right) - K\left(\frac{\boldsymbol{\Psi}_2(f_j^*)-\boldsymbol{x}}{h_n}\right)\right| \mathrm{d}\boldsymbol{x} \\
&\quad + |\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 - \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2| \\
&\quad \left. + |\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_j))\|_2 - \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_j^*))\|_2| \right\}. \qquad (40)
\end{aligned}
$$

Under assumption (iv), following Theorems 3.1 and 3.2 and (37), we have

$$
\begin{aligned}
&|\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 - \|\nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2| \\
&\quad \leq \|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i)) - \nabla\lambda_n(\boldsymbol{\Psi}_2(f_i^*))\|_2 + \|\nabla\lambda_n(\boldsymbol{\Psi}_2(f_i^*)) - \nabla\widehat{\lambda}_n(\boldsymbol{\Psi}_2(f_i^*))\|_2 \\
&\quad \leq O(1)\|\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_2(f_i^*)\|_2 + \|\widehat{\lambda}_n - \lambda_n\|_{1,\max}^* \\
&\quad \leq O(1)\,\mathsf{Haus}(\mathcal{L}_n, \widehat{\mathcal{L}}_n) + \|\widehat{\lambda}_n - \lambda_n\|_{1,\max}^*
\end{aligned}
$$

$$= O\left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2}}}\right) \tag{41}$$

holds uniformly over $i = 1, \dots, N(\epsilon, n)$. It follows from assumption (ii) and (37) that

$$\begin{aligned}
& h_n^{-d} \int_{[\mathbf{0},\boldsymbol{\tau}_0]} \left| K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right) - K\left(\frac{\boldsymbol{\Psi}_2(f_i^*) - \boldsymbol{x}}{h_n}\right)\right| \mathrm{d}\boldsymbol{x} \\
&= \int_{[-\mathbf{1},\mathbf{1}]} \left| K(\boldsymbol{t}) - K\left(\boldsymbol{t} + \frac{\boldsymbol{\Psi}_2(f_i^*) - \boldsymbol{\Psi}_1(f_i)}{h_n}\right)\right| \mathrm{d}\boldsymbol{t} \\
&\le O(1)\|\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_2(f_i^*)\|_2 / h_n \\
&= O\left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2}}}\right)
\end{aligned} \tag{42}$$

for all $i = 1, \dots, N(\epsilon, n)$. Combing (40), (41) and (42), we obtain that

$$\delta_{12} = \sup_{1 \le i,j \le N(\epsilon,n)} |\Sigma_1^{i,j} - \Sigma_2^{i,j}| \le c_{36} \sqrt{\frac{|\log h_n|}{nh_n^{d+2}}}$$

for some constant c_{36}. Immediately,

$$\begin{aligned}
\zeta_{n2}^{(1)} + \zeta_{n2}^{(2)} &\le c_{35}(c_{36})^{1/3} \left(\sqrt{\frac{|\log h_n|}{nh_n^{d+2}}}\right)^{1/3} \left\{\log\left(\frac{N(\epsilon, n)\sqrt{nh_n^{d+2}}}{c_{36}\sqrt{|\log h_n|}}\right)\right\}^{2/3} \\
&\quad + (c_{34}c_{31} + c_{33}) \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8} + \kappa_n.
\end{aligned} \tag{43}$$

We now consider κ_n. By the definition of $\breve{\mathscr{F}}$, for any $f \in \mathscr{F}$, there exist $f_i \in \breve{\mathscr{F}}$ and some constant c_{37}, not depending on the choices of f and f_i, such that

$$\begin{aligned}
&\int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d} \left\{\frac{1}{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2} K\left(\frac{\boldsymbol{\Psi}_1(f) - \boldsymbol{x}}{h_n}\right) - \frac{1}{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2} K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{x} \\
&\quad \le c_{37}\epsilon^2,
\end{aligned}$$

from which we have $\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2 / h_n$ is small when ϵ is small. On the other hand, by assumption (ii), there exist some domain $\mathcal{D}_i$ lies in $[-\mathbf{1}, \mathbf{1}]$ and some positive constant c_{38}, not depending on f and f_i, such that $|K(\boldsymbol{t}) - K(\boldsymbol{t} + \frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)}{h_n})| > c_{38}\|\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)}{h_n}\|_2$ for all $\boldsymbol{t} \in \mathcal{D}_i$ using the directional derivative along the particular direction of $\{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)\}/h_n$, which excludes the degenerated case of the kernel function taking a constant. As a consequence,

$$\begin{aligned}
&\int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d} \left\{K\left(\frac{\boldsymbol{\Psi}_1(f) - \boldsymbol{x}}{h_n}\right) - K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{x} \\
&= \int_{[-\mathbf{1},\mathbf{1}]} \left\{K(\boldsymbol{t}) - K\left(\boldsymbol{t} + \frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{t}
\end{aligned}$$

$$\geq \int_{\mathcal{D}_i} \left\{ K(\boldsymbol{t}) - K\left(\boldsymbol{t} + \frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{t}$$
$$\geq c_{39}\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2^2/h_n^2 \tag{44}$$

for some positive constant c_{39}. Actually, (44) still holds for constant kernel function. Note that there exists some constant c_{40} such that

$$\begin{aligned} |\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 - \|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2| &\leq \|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i)) - \nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2 \\ &\leq O(1)\|\lambda_n\|_{2,\max}^*\|\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)\|_2 \\ &\leq c_{40}\|\boldsymbol{\Psi}_1(f_i) - \boldsymbol{\Psi}_1(f)\|_2. \end{aligned}$$

Consequently, following assumptions (ii), (iv) and (44) and some basic calculations, we have

$$\begin{aligned} c_{37}\epsilon^2 \geq &\int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d}\left\{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2^{-1}K\left(\frac{\boldsymbol{\Psi}_1(f) - \boldsymbol{x}}{h_n}\right)\right. \\ &\left. -\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2^{-1}K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{x} \\ \geq\ & c_{41}\int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d}\left\{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 K\left(\frac{\boldsymbol{\Psi}_1(f) - \boldsymbol{x}}{h_n}\right)\right. \\ &\left. -\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2 K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{x} \\ =\ & c_{41}\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2^2 \int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d}\left\{K\left(\frac{\boldsymbol{\Psi}_1(f) - \boldsymbol{x}}{h_n}\right) - K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{x} \\ &+ \{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 - \|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2\}^2 \int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d}\left\{K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\}^2 \mathrm{d}\boldsymbol{x} \\ &+ 2\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2\{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2 - \|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2\} \\ &\times \int_{[\mathbf{0},\boldsymbol{\tau}_0]} h_n^{-d}\left\{K\left(\frac{\boldsymbol{\Psi}_1(f) - \boldsymbol{x}}{h_n}\right) - K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right)\right\} K\left(\frac{\boldsymbol{\Psi}_1(f_i) - \boldsymbol{x}}{h_n}\right) \mathrm{d}\boldsymbol{x} \\ \geq\ & c_{42}\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2^2/h_n^2 - c_{43}\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2^2/h_n \\ \geq\ & c_{44}\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2^2/h_n^2 \end{aligned}$$

for some positive constants c_{41}, c_{42}, c_{43} and c_{44}, which leads to

$$\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2 \leq (c_{37}c_{44}^{-1})^{1/2}h_n\epsilon.$$

Therefore, under assumptions (ii) and (iv), when $\|\boldsymbol{\Psi}_1(f) - \boldsymbol{\Psi}_1(f_i)\|_2 \leq (c_{37}c_{44}^{-1})^{1/2}h_n\epsilon$, we have

$$\begin{aligned} &\left|\frac{|\widehat{\lambda}_n(\boldsymbol{\Psi}_1(f)) - \lambda_n(\boldsymbol{\Psi}_1(f))|}{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f))\|_2} - \frac{|\widehat{\lambda}_n(\Xi_1(f_i)) - \lambda_n(\boldsymbol{\Psi}_1(f_i))|}{\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2}\right| \\ &\quad \leq O(1)|\|\nabla\lambda_n(\boldsymbol{\Psi}_1(f_i))\|_2|\widehat{\lambda}_n(\boldsymbol{\Psi}_1(f)) - \lambda_n(\boldsymbol{\Psi}_1(f))| \end{aligned}$$

$$
\begin{aligned}
&- \|\nabla\lambda_n(\mathbf{\Psi}_1(f))\|_2|\widehat{\lambda}_n(\mathbf{\Psi}_1(f_i)) - \lambda_n(\mathbf{\Psi}_1(f_i))|\,| \\
&\le O(1)\{|\|\nabla\lambda_n(\mathbf{\Psi}_1(f))\|_2 - \|\nabla\lambda_n(\mathbf{\Psi}_1(f_i))\|_2| \\
&\quad + |\,|\widehat{\lambda}_n(\mathbf{\Psi}_1(f)) - \lambda_n(\mathbf{\Psi}_1(f))| - |\widehat{\lambda}_n(\mathbf{\Psi}_1(f_i)) - \lambda_n(\mathbf{\Psi}_1(f_i))|\,|\} \\
&\le O(1)\{\|\nabla\lambda_n(\mathbf{\Psi}_1(f)) - \nabla\lambda_n(\mathbf{\Psi}_1(f_i))\|_2 + |\widehat{\lambda}_n(\mathbf{\Psi}_1(f)) - \widehat{\lambda}_n(\mathbf{\Psi}_1(f_i))| \\
&\quad + |\lambda_n(\mathbf{\Psi}_1(f)) - \lambda_n(\mathbf{\Psi}_1(f_i))|\} \\
&\le O(1)(\|\lambda_n\|^*_{2,\max} + \|\widehat{\lambda}_n\|^*_{1,\max} + \|\lambda_n\|^*_{1,\max})\|\mathbf{\Psi}_1(f) - \mathbf{\Psi}_1(f_i)\|_2 \\
&\le c_{45}h_n\epsilon
\end{aligned}
$$

for some constant c_{45}. Immediately,

$$
\sup_{f\in\mathscr{F}} \frac{|\widehat{\lambda}_n(\mathbf{\Psi}_1(f)) - \lambda_n(\mathbf{\Psi}_1(f))|}{\|\nabla\lambda_n(\mathbf{\Psi}_1(f))\|_2} \le \max_{1\le i\le N(\epsilon,n)} \frac{|\widehat{\lambda}_n(\mathbf{\Psi}_1(f_i)) - \lambda_n(\mathbf{\Psi}_1(f_i))|}{\|\nabla\lambda_n(\mathbf{\Psi}_1(f_i))\|_2} + c_{45}h_n\epsilon,
$$

which implies that

$$
\begin{aligned}
&\left| \sup_{f\in\mathscr{F}} \sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\mathbf{\Psi}_1(f)) - \lambda_n(\mathbf{\Psi}_1(f))|}{\|\nabla\lambda_n(\mathbf{\Psi}_1(f))\|_2} - \sup_{f\in\breve{\mathscr{F}}} \sqrt{nh_n^d}\frac{|\widehat{\lambda}_n(\mathbf{\Psi}_1(f)) - \lambda_n(\mathbf{\Psi}_1(f))|}{\|\nabla\lambda_n(\mathbf{\Psi}_1(f))\|_2} \right| \\
&\le c_{45}\sqrt{nh_n^{d+2}}\epsilon.
\end{aligned}
$$

By setting $\epsilon = \frac{c_{31}\log^{3/8} n}{6c_{45}(nh_n)^{1/8}(nh_n^{d+2})^{1/2}}$, it is straightforward to show that $\kappa_n = 0$. For such chosen ϵ, following assumption (iii) and Remark 1, there exists some constant c_{46} such that

$$
\frac{N(\epsilon, n)\sqrt{nh_n^{d+2}}}{c_{36}\sqrt{|\log h_n|}} \le n^{c_{46}}.
$$

Hence, (43) can be deduced as

$$
\begin{aligned}
\zeta_{n2}^{(1)} + \zeta_{n2}^{(2)} &\le c_{35}(c_{36})^{1/3}(c_{46})^{2/3}\left(\frac{|\log h_n|\log^4 n}{nh_n^{d+2}}\right)^{1/6} + (c_{34}c_{31} + c_{33})\left(\frac{\log^7 n}{nh_n^d}\right)^{1/8} \\
&= O\left(\left(\frac{\log^5 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right) \qquad (45)
\end{aligned}
$$

by utilising Remark 3.1.

Employing analogous arguments, we have

$$
\zeta_{n2}^{(3)} + \zeta_{n2}^{(4)} = O\left(\left(\frac{\log^5 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right). \qquad (46)
$$

Following (38), (45) and (46), we can conclude

$$
\zeta_{n2} = O\left(\left(\frac{\log^5 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right). \qquad (47)
$$

Under assumption (iii), it follows from (33), (34), (35), (47) and Remark 3.1 that

$$\sup_t |\mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n^*, \widehat{\mathcal{L}}_n) \le t|\mathcal{O}_1,\ldots,\mathcal{O}_n) - \mathbb{P}(\sqrt{nh_n^d}\,\mathsf{Haus}(\widehat{\mathcal{L}}_n, \mathcal{L}_n) \le t)|$$

$$= O\left(\left(\frac{\log^5 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^3 n}{nh_n^{d+2}}\right)^{1/2} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right)$$

$$= O\left(\left(\frac{\log^7 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right),$$

which implies that

$$\mathbb{P}(\mathcal{L}_n \subset \widehat{\mathcal{L}}_n \oplus w_{1-\alpha}^*) \ge 1 - \alpha + O\left(\left(\frac{\log^7 n}{nh_n^{d+2}}\right)^{1/6} + \left(\frac{\log^7 n}{nh_n^d}\right)^{1/8}\right)$$

using $\mathbb{P}(\mathcal{L}_n \subset \widehat{\mathcal{L}}_n \oplus w_{1-\alpha}) \ge 1 - \alpha$. This completes the proof of Theorem 3.5. ■

Acknowledgements

We thank the Associate Editor and two referees for their constructive comments and helpful suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

G. Mao, Y. Liu and Y. Wu's research were partly supported by the National Natural Science Foundation of China [grant numbers 12101256, 11971362, 12071483]; and Q. Yang's research was partly supported by the National Social Science Foundation of China [grant number 20BTJ039].

ORCID

Guangcai Mao http://orcid.org/0000-0002-4049-1065
Yuanshan Wu http://orcid.org/0000-0002-2121-8952

References

Burke, M.D. (1988), 'Estimation of a Bivariate Distribution Function Under Random Censorship', *Biometrika*, 75, 379–382.

Calonico, S., Cattaneo, M.D., and Farrell, M.H. (2017), 'On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference', *Journal of the American Statistical Association*, 113, 767–779.

Campbell, G., and Földes, A. (1982), 'Large-sample Properties of Nonparametric Bivariate Estimators with Censored Data', in *Nonparametric Statistical Inference* (eds. B.V. Gnedenko, M.L. Puri and I. Vincze), pp. 103–122. Amsterdam: North-Holland.

Chen, Y.-C., Genovese, C.R., and Wasserman, L. (2017), 'Density Level Sets: Asymptotics, Inference, and Visualization', *Journal of the American Statistical Association*, 112, 1684–1696.

Chernozhukov, V., Hong, H., and Tamer, E. (2007), 'Estimation and Confidence Regions for Parameter Sets in Econometric Models', *Econometrica*, 75, 1243–1284.
Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), 'Comparison and Anti-concentration Bounds for Maxima of Gaussian Random Vectors', *Probability Theory and Related Fields*, 162, 47–70.
Chernozhukov, V., Chetverikov, D., and Kato, K. (2014), 'Gaussian Approximation of Suprema of Empirical Processes', *The Annals of Statistics*, 42, 1564–1597.
Cuevas, A., González-Manteiga, W., and Rodríguez-Casal, A. (2006), 'Plug-in Estimation of General Level Sets', *Australian and New Zealand Journal of Statistics*, 48, 7–19.
Dabrowska, D.M. (1988), 'Kaplan–Meier Estimate on the Plane', *The Annals of Statistics*, 16, 1475–1489.
Dabrowska, D.M. (1989), 'Kaplan–Meier Estimate on the Plane: Weak Convergence, LIL, and the Bootstrap', *Journal of Multivariate Analysis*, 29, 308–325.
Diehl, S., and Stute, W. (1988), 'Kernel Density and Hazard Function Estimation in the Presence of Censoring', *Journal of Multivariate Analysis*, 25, 299–310.
Gill, R.D., van der Laan, M.J., and Wellner, J.A. (1995), 'Inefficient Estimators of the Bivariate Survival Function for Three Models', *Annales de l'Institut Henri Poincaré*, 31, 545–597.
Gilstein, C.Z., and Leamer, E.E. (1983), 'Robust Sets of Regression Estimates', *Econometrica*, 51, 321–333.
Giné, E., and Guillou, A. (2001), 'On Consistency of Kernel Density Estimators for Randomly Censored Data: Rates Holding Uniformly Over Adaptive Intervals', *Annales de l'Institut Henri Poincaré*, 37, 503–522.
Giné, E., and Guillou, A. (2002), 'Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators', *Annales de l'Institut Henri Poincaré*, 38, 907–921.
Jankowski, H., Ji, X., and Stanberry, L. (2014), 'A Random Set Approach to Confidence Regions with Applications to the Effective Dose with Combinations of Agents', *Statistics in Medicine*, 33, 4266–4278.
Laloë, T., and Servien, R. (2013), 'Nonparametric Estimation of Regression Level Sets Using Kernel Plug-in Estimator', *Journal of the Korean Statistical Society*, 42, 301–311.
Lo, S.H., Mack, Y.P., and Wang, J.L. (1989), 'Density and Hazard Rate Estimation for Censored Data Via Strong Representation of the Kaplan–Meier Estimator', *Probability Theory and Related Fields*, 80, 461–473.
Mammen, E., and Polonik, W. (2013), 'Confidence Regions for Level Sets', *Journal of Multivariate Analysis*, 122, 202–214.
Patil, P.N. (1993), 'On the Least Squares Cross-validation Bandwidth in Hazard Rate Estimation', *The Annals of Statistics*, 21, 1792–1810.
Polonik, W. (1995), 'Measuring Mass Concentrations and Estimating Density Contour Clusters: An Excess Mass Approach', *The Annals of Statistics*, 23, 855–881.
Prentice, R.L., and Cai, J. (1992), 'Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data', *Biometrika*, 79, 495–512.
Pruitt, R.C. (1991), 'Strong Consistency of Self-consistent Estimators: General Theory and an Application to Bivariate Survival Analysis', Technical Report 543, University of Minnesota, School of Statistics.
Ramlau-Hansen, H. (1983), 'Smoothing Counting Process Intensities by Means of Kernel Functions', *The Annals of Statistics*, 11, 453–466.
Rinaldo, A., and Wasserman, L. (2010), 'Generalized Density Clustering', *The Annals of Statistics*, 38, 2678–2722.
Romano, J.P., and Shaikh, A.M. (2010), 'Inference for the Identified Set in Partially Identified Econometric Models', *Econometrica*, 78, 169–211.
Singh, A., Scott, C., and Nowak, R. (2009), 'Adaptive Hausdorff Estimation of Density Level Sets', *The Annals of Statistics*, 37, 2760–2782.
Tanner, M.A., and Wong, W.H. (1983), 'The Estimation of the Hazard Function From Randomly Censored Data by the Kernel Method', *The Annals of Statistics*, 11, 989–993.

Tsybakov, A.B. (1997), 'On Nonparametric Estimation of Density Level Sets', *The Annals of Statistics*, 25, 948–969.
Walther, G. (1997), 'Granulometric Smoothing', *The Annals of Statistics*, 25, 2273–2299.
Watson, G.S., and Leadbetter, M.R. (1964), 'Hazard Analysis. I', *Biometrika*, 51, 175–184.
Yandell, B.S. (1983), 'Nonparametric Inference for Rates with Censored Survival Data', *The Annals of Statistics*, 11, 1119–1135.