ON LOCAL POLYNOMIAL ESTIMATION OF HAZARD RATES AND THEIR DERIVATIVES UNDER RANDOM CENSORING

JIANCHENG JIANG
Peking University

KJELL DOKSUM
University of California at Berkeley

The Dirac function is used to define an empirical hazard rate $\lambda_n(\cdot)$ whose integral up to time t equals to the Nelson-Aalen estimator. This empirical hazard rate exists only in space of Schwartz distributions, so we introduce a local polynomial approximation to $\lambda_n(\cdot)$ which provides estimators of the hazard rate and its derivatives. Consistency and joint asymptotic normality of the local polynomial estimators are established. The estimators have favorable properties similar to those of local polynomial regression estimators, that is, the hazard rate estimator is boundary adaptive and under certain smoothness conditions the rate of convergence can be made arbitrary close to root n. The estimator is boundary corrected even if a local constant smoother is employed. Asymptotic expressions for the mean squared errors (MSE's) are obtained and used in bandwidth selection. A data-driven local bandwidth selection rule is proposed and is illustrated on the Stanford heart transplant data. We use Monte Carlo methods to show that the proposed estimator compares favorably with the Müller–Wang estimator.

1. Introduction

Assume that $T_1, \ldots, T_n$ are i.i.d. lifetimes (that is, nonnegative random variables) with distribution function F, and that $C_1, \ldots, C_n$ are i.i.d. censoring times with distribution function G. The C_i, T_i are assumed to be independent, and the actual observations are (X_i, δ_i), for $i = 1, \ldots, n$, where $X_i = \min(T_i, C_i)$ and $\delta_i = I(X_i = T_i)$ is an indicator of the censoring status of X_i.

Let L denote the distribution function of X_i, then $\bar{L} = \overline{F}\,\overline{G}$, where for any distribution function E, $\overline{E} = 1 - E$ is the corresponding survival function. Let $\Lambda(x) = -\log(\overline{F}(x))$ be the cumulative hazard function. We consider the problem of estimating $\lambda(x) = \Lambda'(x) = f(x)/\overline{F}(x)$ and $\lambda^{(k)}(x)$ for $k \geq 0$ on the interval $[0, T]$, where $T < T^* \equiv \inf\{x : L(x) = 1\}$.

Ramlau-Hausen (1983), Tanner and Wong (1983), and Yandell (1983) studied the asymptotic properties of kernel estimators of hazard functions based on the idea of convolution. Müller and Wang (1990) considered local bandwidth choice for convolution-type kernel estimators with fixed higher order kernels, and Müller and Wang (1994) proposed to estimate hazard functions with varying kernels and data-adaptive bandwidths in order to remove boundary effects. Hess et al. (1999) reviewed various kernel-based

Keywords and phrases: Censoring; data-driven local bandwidth selection; Dirac function; hazard rate; local polynomial..

AMS subject classifications: Primary 62G07; Secondary 62G20, 62E25.

estimators for hazard functions and advocated the use of boundary correction and locally optimal bandwidths.

Here we consider local polynomial (LP) estimators of hazard rate functions and their derivatives that are approximations to Dirac derivatives of the Nelson-Aalen estimator and show that the LP estimators share some favorable properties with local polynomial regression estimators. In particular, the LP estimators can reduce the bias according to the degree of the polynomial without increasing the variance and automatically correct the left boundary effect. Moreover, the finite sample bias of the LP estimators is zero when the estimated hazard rate is a polynomial up to order p, where p is the order of the polynomial used in smoothing (see (2.3) below). The pointwise asymptotic normality of the LP estimator enables one to find the asymptotically optimal variable bandwidth, and allows one to develop a data-driven optimal local bandwidth selector by using the ideas of Fan and Gijbels (1995). In this paper, we present a simple data-driven method for choosing the local bandwidth.

The outline of this paper is as follows. In Section 2, we introduce the LP estimators. Section 3 concentrates on the asymptotic properties of the proposed estimators, including pointwise strong consistency and joint asymptotic normality. In Section 4, the data-driven local bandwidth selection rule is proposed. Numerical illustration is given in Section 5. Technical proofs are given in the appendix.

2. Estimation

In order to introduce the estimators, we will use the following notation:

(1) $L_1(x) = P(X_i \leq x, \delta_i = 1)$, the subdistribution function for the uncensored observations.

(2) $L_{1n}(x) = \sum_{i=1}^n I(X_i \leq x, \delta_i = 1)/(n+1)$, the modified empirical distribution function of $L_1(x)$.

(3) $L_n(x) = \sum_{i=1}^n I(X_i \leq x)/(n+1)$, the modified empirical distribution function of $L(x)$.

Note that $\lambda(x) = L_1'(x)/\bar{L}(x)$. We will use the Nelson-Aalen's estimator of $\Lambda(x)$,

$$\Lambda_n(x) = \int_0^x (1 - L_n(u))^{-1}\, dL_{1n}(u) = \sum_{i:X_{(i)} \leq x} \delta_{(i)}/(n-i+1), \tag{2.1}$$

where $X_{(i)}$ are the order statistics of X_i, and $\delta_{(i)}$ is the concomitant of $X_{(i)}$, i.e., $\delta_{(i)}$ and $X_{(i)}$ are from the same observation.

For a given point $x_0 \in (0, T)$, the following assumptions and notations are needed.

(A1) The hazard function $\lambda(x)$ has a continuous $(p+2)th$ derivative at the point x_0.

(A2) The sequence of bandwidths b_n tends to zero such that $nb_n \to +\infty$ as $n \to +\infty$. Let $\mathbf{B} = \mathrm{diag}(1, b_n, \ldots, b_n^p)$.

(A3) $L(x)$ is continuous at the point x_0.

(A4) The kernel function K is a continuous function of bounded variation and with bounded support $[-1,1]$, say. Let $s_\ell = \int_{-1}^{1} K(u)u^\ell\, du$, $v_\ell = \int_{-1}^{1} u^\ell K^2(u)\, du$, $\mathbf{c}_p = (s_{p+1}, \ldots, s_{2p+1})^T$, $\mathbf{c}_{p+1} = (s_{p+2}, \ldots, s_{2p+2})^T$, $\mathbf{S} = (s_{i+j})$ and $\mathbf{V}^* = (v_{i+j})$, $(0 \le i \le p;\ 0 \le j \le p)$.

Müller and Wang (1990, 1994) considered the following kernel estimator, which is a convolution of the Nelson-Aalen estimator Λ_n with an appropriate kernel function K_ν:

$$\begin{aligned} \hat{\lambda}^{(\nu)}(x) &= \frac{1}{b_n^{\nu+1}} \int K_\nu\left(\frac{x-u}{b_n}\right) d\Lambda_n(u) \\ &= \frac{1}{b_n^{\nu+1}} \sum_{i=1}^{n} K_\nu\left(\frac{x - X_{(i)}}{b_n}\right) \frac{\delta_{(i)}}{n-i+1}, \end{aligned} \tag{2.2}$$

where K_ν is a kernel of order (ν, k) with $k > \nu$ (see (2.5) in Müller and Wang, 1990). For the estimation of derivatives or reduction of bias, the estimator needs higher order kernels, which can lead to a negative hazard rate estimator. The practical advantages of using higher order kernels can be quite small for moderate sample sizes as demonstrated in Marron and Wand (1992). When estimating at a point x near 0 or T, most kernel estimators in the density estimation and regression setting will encounter boundary effects. The estimator (2.2) suffers from boundary effects near the endpoints of the support of the hazard rates. Müller and Wang (1994) proposed to solve the problem by employing boundary kernels and a data-adaptive varying bandwidth selection procedure. Hall and Wehrly (1991) studied a geometrical method for removing edge effects from kernel-type nonparametric regression estimators, which may be useful in density estimation. Here we introduce a simple and intuitive approach to the problem, which does not employ higher order kernels or boundary kernels while automatically correcting the boundary effects.

Let us consider the following optimization problem: for $p = 0, 1, 2, \ldots$,

$$\min_{a_j} \int_{u \ge 0} \frac{1}{b_n} K\left(\frac{u-x}{b_n}\right) \left[\lambda(u) - \sum_{j=0}^{p} a_j (u-x)^j\right]^2 du. \tag{2.3}$$

By Taylor expansion, the solution of the optimization problem, denoted by $\mathbf{a}^*(x) \equiv (a_0^*, \ldots, a_p^*)^T$, approximates $\mathbf{a}(x) \equiv (\lambda(x), \ldots, \lambda^{(p)}(x)/p!)^T$.

Note that the Nelson-Aalen's estimator $\Lambda_n(x)$ is the empirical estimator of $\Lambda(x)$, thus we can define the following *generalized empirical hazard rate*:

$$\lambda_n(x) = \sum_{i=1}^{n} \frac{\delta_{(i)}}{n-i+1} D(x - X_{(i)}), \tag{2.4}$$

where $D(x)$ is the Dirac function with the following property:

$$\int g(u)D(u-x)\,du = g(x)$$

for any integrable function $g(x)$. Then $\int_0^x \lambda_n(t)\,dt = \Lambda_n(x)$, however, $\lambda_n(\cdot)$ exists only in space of Schwartz distributions and is not computable, which is why we call λ_n the *generalized empirical hazard rate.*

Because integrals involving $\lambda_n(\cdot)$ exist we can obtain computable estimators of $\mathbf{a}(x)$ for x fixed by using an empirical version of (2.3). That is, we define the LP estimator of $\mathbf{a}(x)$ as

$$\begin{aligned} \hat{\mathbf{a}}(x) &\equiv (\hat{a}_0, \dots, \hat{a}_p)^T \\ &= \arg\min_{a_j} \int_{u \geq 0} \frac{1}{b_n} K\left(\frac{u-x}{b_n}\right) \left[\lambda_n(u) - \sum_{j=0}^{p} a_j (u-x)^j\right]^2 du. \end{aligned} \tag{2.5}$$

In estimation of distribution and density functions without censoring, similar ideas have been used. Lejeune and Sarda (1992) considered estimation of the distribution function by local linear fitting to the empirical distribution F_n. Jones (1993) considered a locally linear estimator and established its link with the generalized jackknife boundary correction for $p = 1$ by smoothing the generalized empirical density function. Nielsen and Tanggaard (2001) used the counting process approach to construct estimates similar to our locally constant and locally linear estimates.They also provide a number of useful references. Here we will consider the local polynomial estimation of hazard functions and their derivatives, $\mathbf{a}(x)$, in the case of censoring. We will show that the above LP estimators for hazard rates automatically correct the boundary effect even if a local constant smoother is used.

Taking the derivative with respect to the $a's$ of the integral in (2.5), we obtain the LP estimator $\hat{\mathbf{a}}(x)$ as the solution to the linear equations: for $\ell = 0, \dots, p$,

$$\begin{aligned} &\sum_{i=1}^{n} \frac{1}{b_n} K\left(\frac{X_{(i)} - x}{b_n}\right) (X_{(i)} - x)^{\ell} \frac{\delta_{(i)}}{n-i+1} \\ &\qquad = \sum_{i=0}^{p} a_i \int_{u \geq 0} (u-x)^{i+\ell} \frac{1}{b_n} K\left(\frac{u-x}{b_n}\right) du. \end{aligned} \tag{2.6}$$

It follows that the LP estimator at an interior ($x_0 \in [b_n, T)$) point satisfies the following closed form

$$\mathbf{B}\hat{\mathbf{a}}(x_0) = \mathbf{S}^{-1}\mathbf{S_n}(x_0), \tag{2.7}$$

where $\mathbf{S_n}(x_0) = (S_{n0}(x_0), \cdots, S_{np}(x_0))^T$, and

$$S_{n\ell}(x_0) = \sum_{i=1}^{n} \frac{1}{b_n} K\left(\frac{X_{(i)} - x_0}{b_n}\right)\left(\frac{X_{(i)} - x_0}{b_n}\right)^{\ell} \frac{\delta_{(i)}}{n-i+1}. \tag{2.8}$$

When x_0 is an interior point, $p = 0$ and $s_0 = 1$, (2.8) gives

$$\hat{a}_0 = \sum_{i=1}^{n} \frac{1}{b_n} K(\frac{X_{(i)} - x_0}{b_n}) \frac{\delta_{(i)}}{n-i+1},$$

which is the same as the estimator of Müller and Wang (1990, 1994) for the hazard rate, so that Müller and Wang's estimator with $v = 0$ at an interior point x_0 coincides with the LP estimators with $p = 0$. However, this equivalence does not hold for boundary ($x \in [0, b_n)$) points.

We will show in the next section that the LP estimator shares nice properties with the local polynomial regression estimator, in particular, the estimator automatically corrects the left boundary effect, which contrasts with the results for other hazard rate estimators. Asymptotically, when $L(T) < 1$ and $b_n \to 0$, we will not encounter boundary effects at T.

3. Asymptotic properties

In this section, we will establish the consistency and joint asymptotic normality of the local polynomial estimators.

Theorem 3.1. *Under conditions* (A1)–(A4),

$$\mathbf{B}\big(\hat{\mathbf{a}}(x_0) - \mathbf{a}(x_0)\big) \xrightarrow{\text{a.s.}} 0, \quad n \to \infty.$$

Theorem 3.2. *Under conditions* (A1)–(A4),

$$\begin{aligned}\sqrt{nb_n}\Big\{\mathbf{B}[\hat{\mathbf{a}}(x_0) - \mathbf{a}(x_0)] &- \frac{\lambda^{(p+1)}(x_0)b_n^{p+1}}{(p+1)!}\mathbf{S}^{-1}\mathbf{c_p} \\ &- \frac{\lambda^{(p+2)}(x_0)b_n^{p+2}}{(p+2)!}\mathbf{S}^{-1}\mathbf{c_{p+1}} + o(b_n^{p+2})\Big\} \\ &\xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\frac{\lambda(x_0)}{\bar{L}(x_0)}\Big).\end{aligned} \tag{3.1}$$

Remark 3.1. When estimating a hazard rate which is a polynomial of order p on an interval, the finite sample bias of the LP estimators on the interval is zero (see the proof of Theorem 3.2). This contrasts with the methods of Müller and Wang (1990, 1994) based on higher order kernels, for which the respective zero bias only holds true asymptotically.

Remark 3.2. Consider the left edge effect on the estimator. A convenient mathematical formulation of the edge effect problem is given by Gasser and Müller (1979). Assume that we estimate $\mathbf{a}(x)$ at $x_n = db_n$ in the left boundary region for some positive constant $d \in [0, 1]$. Then similar to (2.7), $\hat{\mathbf{a}}(x_n)$ in (2.6) has the following closed form, for $p = 0, 1, 2, \ldots,$

$$\mathbf{B}\hat{\mathbf{a}}(x_n) = \mathbf{S}_d^{-1}\mathbf{S}_n(x_n), \tag{3.2}$$

where $\mathbf{S}_d$ defined as $\mathbf{S}$ but with s_i replaced by $s_{i,d} = \int_{-d}^{1} u^i K(u)\,du$. Let $v_{i,d} = \int_{-d}^{1} u^i K^2(u)\,du$. Then the joint asymptotic normality (3.1) continues to hold with $\mathbf{c}_p$, $\mathbf{c}_{p+1}$, $\mathbf{S}$, and $\mathbf{V}^*$ replaced by $\mathbf{c}_{p,d}$, $\mathbf{c}_{p+1,d}$, $\mathbf{S}_d$, and $\mathbf{V}_d^*$, respectively, where $\mathbf{c}_{p,d} = (s_{p+1,d}, \ldots, s_{2p+1,d})^T$, $\mathbf{c}_{p+1,d} = (s_{p+2,d}, \ldots, s_{2p+2,d})^T$, and $\mathbf{V}_d^* = (v_{i+j,d})$ is $(p+1) \times (p+1)$ matrices. This property is similar to that of local polynomial regression estimation, which is not shared by other kernel estimators of hazard rates (see for example Hess et al., 1999). The LP estimators are automatically boundary adaptive in the sense of Fan and Gijbels (1996). Note that the above property holds even for $p = 0$, which contrasts with the cases of local polynomial regression.

Note that when x_0 is an interior point and $K(\cdot)$ is symmetric, some of the entries in $\mathbf{S}^{-1}\mathbf{c}_p$ are zero and Theorem 3.2 is not very informative. We now take a closer look at this issue. Let $\hat{\lambda}^{(k)}(x) = k!\hat{a}_k(x)$, and $\mathbf{e}_k = (0, \ldots, 1, \ldots, 0)^T$, where $\mathbf{e}_k$ has one in the $(k+1)$th component and zeros in the others. Then

Theorem 3.3. *Suppose* $K(\cdot)$ *is symmetric. Under conditions* (A1)–(A4),

(i) *For* $p-k$ *odd*

$$\sqrt{nb_n^{2k+1}}\left\{\frac{1}{k!}[\hat{\lambda}^{(k)}(x_0) - \lambda^{(k)}(x_0)] - \frac{\lambda^{(p+1)}(x_0)b_n^{p+1-k}}{(p+1)!}\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c_p}(1+o(1))\right\}$$
$$\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\mathbf{e}_k\frac{\lambda(x_0)}{\bar{L}(x_0)}\right).$$

(ii) *For* $p-k$ *even*

$$\sqrt{nb_n^{2k+1}}\left\{\frac{1}{k!}[\hat{\lambda}^{(k)}(x_0) - \lambda^{(k)}(x_0)] - \frac{\lambda^{(p+2)}(x_0)b_n^{p+2-k}}{(p+2)!}\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c_{p+1}}(1+o(1))\right\}$$
$$\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\mathbf{e}_k\frac{\lambda(x_0)}{\bar{L}(x_0)}\right).$$

Remark 3.3. From Theorem 3.3, the asymptotic mean squared error MSE for estimating $\lambda^{(k)}(x_0)/k!$ $(k = 0, \ldots, p)$ can be defined as

$$\mathrm{MSE}_k(b_n, x_0) = \begin{cases} b_n^{2(p+1-k)}[\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c}_\mathbf{p}\frac{\lambda^{(p+1)}(x_0)}{(p+1)!}]^2 + \frac{1}{nb_n^{2k+1}}\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\mathbf{e}_k\frac{\lambda(x_0)}{\bar{L}(x_0)}, & \text{if } p-k \text{ is odd;} \\ b_n^{2(p+2-k)}[\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c}_{p+1}\frac{\lambda^{(p+2)}(x_0)}{(p+2)!}]^2 + \frac{1}{nb_n^{2k+1}}\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\mathbf{e}_k\frac{\lambda(x_0)}{\bar{L}(x_0)}, & \text{otherwise.} \end{cases}$$

Therefore, the optimal local bandwidth for estimating the kth derivative of $\lambda(x)$ at x_0, in the sense of minimizing $\mathrm{MSE}_k(b_n, x_0)$, is

$$(3.3)\quad b_{k,opt}(x_0) = \begin{cases} n^{-1/(2p+3)}\left(\frac{[(p+1)!]^2\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\mathbf{e}_k\lambda(x_0)/\bar{L}(x_0)}{2(p+1-k)[\lambda^{(p+1)}(x_0)]^2(\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c}_\mathbf{p})^2}\right)^{1/(2p+3)}, & \text{if } p-k \text{ is odd;} \\ n^{-1/(2p+5)}\left(\frac{[(p+2)!]^2\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}\mathbf{e}_k\lambda(x_0)/\bar{L}(x_0)}{2(p+2-k)[\lambda^{(p+2)}(x_0)]^2(\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c}_{p+1})^2}\right)^{1/(2p+5)} & \text{otherwise.} \end{cases}$$

In parallel to Theorem 3.3, consider the estimation of $\lambda^{(k)}(x)$ on the boundary point $x_n = db_n$ for $d \in [0, 1)$. Since $\mathbf{e}_k^T\mathbf{S}_d^{-1}\mathbf{c}_{p,d}$ does not vanish, we have the following theorem from Remark 3.2:

Theorem 3.4. *Suppose $K(\cdot)$ is symmetric. Under conditions* (A1)-(A4),

$$\sqrt{nb_n^{2k+1}}\left\{\frac{1}{k!}[\hat{\lambda}^{(k)}(x_n) - \lambda^{(k)}(x_n)] - \frac{\lambda^{(p+1)}(0)b_n^{p+1-k}}{(p+1)!}\mathbf{e}_k^T\mathbf{S}_d^{-1}\mathbf{c}_{p,d} + o(b_n^{p+1-k})\right\} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \mathbf{e}_k^T\mathbf{S}_d^{-1}\mathbf{V}_d^*\mathbf{S}_d^{-1}\mathbf{e}_k\frac{\lambda(0)}{\bar{L}(0)}\right).$$

Remark 3.4. From Theorems 3.3 and 3.4, the LP estimator of the hazard rate $\lambda(\cdot)$ (for $p = 1$) is boundary adaptive in the sense that it automatically achieves the same convergence rate $O(n^{-2/5})$ on boundary points as in interior region if a symmetric kernel and the optimal bandwidths $b_{0,opt} = O(n^{-1/5})$ in (3.3) are employed. For $p = 0$, the estimator is also consistent, but the bias at the left boundary is of order $O(b_n)$. This contrasts with the case of local constant regression.

4. Data-driven local bandwidth choice

Local bandwidth choice for the estimation of hazard functions and their derivatives is an important problem, especially in practice where one would like to have a data-driven approach to bandwidth choice. Patil (1993) studied least squares cross-validation bandwidth selection in hazard rate estimation, González-Manteiga et al. (1996) studied smoothed bootstrap selection of the global bandwidth for estimation of the hazard function. Müller and Wang (1994) studied the local bandwidth choice for kernel estimators of the hazard functions. Hess et al. (1999) advocated locally optimal bandwidth estimators with left boundary correction. We will study a data-driven optimal local bandwidth choice for the LP estimator.

By examining the MSE of the LP estimator $\hat{\mathbf{a}}(x_0)$ in the proof of Theorem 3.2, we find, for estimating $a_k(x_0) = \lambda^{(k)}(x_0)/k!$, that the exact bias of the estimator $\hat{a}_k(x_0)$ is

$$B_k(b_n, x_0) = \mathbf{e}_k^T\big(\mathbf{S}^{-1}\boldsymbol{\beta}_n(x_0) - \mathbf{B}\mathbf{a}(x_0)\big), \tag{4.1}$$

where $\boldsymbol{\beta}_n(x_0) = \big(\beta_{n0}(x_0), \ldots, \beta_{np}(x_0)\big)^T$, and

$$\begin{aligned}\beta_{nk}(x_0) &= \int \frac{1}{b_n} K\Big(\frac{u - x_0}{b_n}\Big)\Big(\frac{u - x_0}{b_n}\Big)^k \lambda(u)\, du \\ &= \int K(t) t^k \lambda(x_0 + b_n t)\, dt.\end{aligned} \tag{4.2}$$

The asymptotic variance of $\hat{a}_k(x_0)$ is

$$V_k(b_n, x_0) = \frac{1}{nb_n}\mathbf{e}_k^T \mathbf{S}^{-1}\widetilde{\mathbf{V}}\mathbf{S}^{-1}\mathbf{e}_k, \tag{4.3}$$

where $\widetilde{\mathbf{V}} = (\tilde{v}_{ij})$, and $\tilde{v}_{ij} = \int [K^2(t) t^{i+j}\lambda(x_0 + b_n t)/\bar{L}(x_0 + b_n t)]\, dt$. Then we propose to estimate the MSE of $\hat{\lambda}_k(x_0) = k!\hat{a}_k(x_0)$ via

$$\widehat{\mathrm{MSE}}_k(b_n, x_0) = \widehat{B}_k^2(b_n, x_0) + \widehat{V}_k(b_n, x_0), \tag{4.4}$$

where $\widehat{B}_k(b_n, x_0)$ and $\widehat{V}_k(b_n, x_0)$ are defined similarly to $B_k(b_n, x_0)$ and $V_k(b_n, x_0)$ but with $\lambda(x)$ replaced by a pilot estimator and $L(x)$ replaced by its empirical distribution function. Let $\hat{b}_{k,\mathrm{opt}}(x_0) = \arg\min_b \widehat{\mathrm{MSE}}_k(b, x_0)$.

For an illustration, consider data-driven optimal local bandwidth choice for the LP estimator with $p = 0$. Similar approaches can be developed for $p = 1, 2, \ldots$. We now introduce an algorithm for estimating $\lambda(x)$, which is similar to that in Müller and Wang (1994) but enhanced via incorporating our estimator and a local linear smoother for bandwidths. Similar algorithms can be developed for estimating the derivatives of $\lambda(x)$.

Algorithm for estimating $\boldsymbol{\lambda(x)}$

Step 1 (Pilot estimators of $\boldsymbol{\lambda(x)}$). Choose a kernel, such as the Epanechnikov kernel, and an initial global bandwidth b_0. The choice of the initial bandwidth depends on the specific case. Assume the data are available on

$[0, T]$, then a possible value for b_0 is $T/(8n_u^{1/5})^1$ as recommended by Müller and Wang (1994), where n_u is the number of uncensored observations. The pilot estimators $\hat{\lambda}(x)$ of $\lambda(x)$ are obtained by using $b_n(x) \equiv b_0$ and the LP estimator (2.6).

Step 2 (Minimizing of $\widehat{\mathbf{MSE}}(b_n, x)$). Choose an equispaced grid of $m1$ points $\tilde{x}_i$, $i = 1, \ldots, m1$ between 0 and T. For each of the gridpoints $\tilde{x}_i$ compute $\widehat{\mathrm{MSE}}(b_n, \tilde{x}_i)$ in (4.4) with $k = 0$ and obtain its minimizers $\tilde{b}(\tilde{x}_i)$ on the interval $[b_0/4, 4b_0]$,[2] say.

Step 3 (Smoothing bandwidths). Choose another equispaced grid of $m2$ points x_r, $r = 1, \ldots, m2$, over the interval $[0, T]$ on which the final hazard estimator is desired. Running local linear smoother (Fan and Gijbels, 1996) by employing global bandwidth $\tilde{b}_0 = b_0$ or $2b_0$:[3]

$$\hat{b}(x_r) = \sum_{i=1}^{m1} w_i \tilde{b}(\tilde{x}_i) / \sum_{i=1}^{m1} w_i,$$

where

$$w_i = K\left(\frac{\tilde{x}_i - x_r}{\tilde{b}_0}\right)[L_{n,2} - L_{n,1}(\tilde{x}_i - x_r)]$$

and

$$L_{n,j} = \sum_{i=1}^{m1} K(\tilde{x}_i - x_r)/\tilde{b}_0)(\tilde{x}_i - x_r)^j, \quad \text{for } j = 1, 2.$$

Step 4 (Final hazard function estimators). Using (2.6), obtain the estimators $\hat{\lambda}(x_r)$ by employing the bandwidth $\hat{b}(x_r)$, for $r = 1, \ldots, m2$.

Remark 4.1. The above algorithm may be repeated by using the estimators $\hat{\lambda}(x_r)$ in Step 4 as pilot estimators in Step 1 and running Step 2–Step 4 again. The pilot estimators of $\lambda(x)$ in Step 1 may also be obtained via maximum likelihood if one has a plausible parametric model in mind. The local smoother in Step 3 is employed to yield a stable estimator for the hazard rate.

5. Numerical studies

In this section, we check Remark 3.2 and assess the effectiveness of our method in a finite sample situation through simulation studies. We only

[1]Note that the optimal bandwidth is of order $O(n_u^{-1/5})$. Our experience shows that the suggested $b_0 = T/(8n_u^{1/5})$ accords reasonably well on average with the chosen optimal value.

[2]The interval may be larger or small, but our experience shows it is a viable choice.

[3]Our experience suggests that the final estimators for $\lambda(x)$ is insensitive to this choice.

investigate the performance of the estimator for $p = 0$, based on the following two points: (i) the estimator is boundary corrected for $p = 0$; (ii) there are fewer parameters for $p = 0$ than for $p = 1$ and the simulations for $p = 1$ can be made similarly to those for $p = 0$.

We will use the Epanechnikov kernel

$$K(x) = 0.75(1 - x^2)I(|x| \leq 1)$$

for the LP estimator and its corresponding boundary modification kernel for the M&W estimator (see Müller and Wand, 1994) throughout this section. In our simulation studies, three models were used: Uniform, Weibull, and Bathtub, which have different hazard curve structures and are important in practice. The Stanford heart transplant data set was analyzed via our method. In order to control the amount of censoring in the simulation, we use the proportional censorship model, which means that F and G satisfy the Lehmann model $\overline{G} = \overline{F}^{\eta}$ for some $\eta > 0$ (e.g., see Koziol and Green, 1976; González-Manteiga et al., 1996). This gives a probability of censoring $\xi = \eta/(1+\eta)$. Here $\xi = 0$ corresponds to *no censoring.* Note that the use of boundary correction and locally optimal bandwidths in Müller and Wang (1994) (or M&W's for short) were advocated by Hess et al. (1999). For simple comparison, we report the Monte Carlo results for the LP estimator with $p = 0$ and M&W's estimator based on their data-driven local bandwidths, respectively.

Example 5.1. We generated 400 samples, each of size $n = 200$ from the uniform distribution model: $T_j \overset{iid}{\sim} F(x) = \text{Uniform}[0, 1]$. The proportional censorship models for censoring are $C_j \overset{iid}{\sim} G(x) = 1 - \overline{F}^{\eta}(x)$ for $\eta = \frac{1}{9}$ and $\frac{1}{2}$ so that the probabilities for censoring are $\xi = 1/10$ and $\frac{1}{3}$, respectively, where T_j is independent of C_j. Then the hazard rate is a strictly increasing function on $[0, 1]$ with range $[0, +\infty)$.

Figure 1 presents the median performance of our estimator and M&W's, together with the envelopes formed by pointwise 2.5% and 97.5% sample percentiles, among the 400 simulations under 1/10 and $\frac{1}{3}$ censoring. The LP estimators (left panels) do well on the left boundary and compares favorably with the M&W's estimators (right panels).

Example 5.2. We simulated 400 samples of size $n = 200$ from the Weibull distribution: $T_j \overset{iid}{\sim} 1 - \exp(-\sqrt{t}) \equiv F(t)$, $C_j \overset{iid}{\sim} 1 - \overline{F}^{\eta}(t)$ (for $\eta = \frac{1}{9}$ and $\frac{1}{2}$), and T_j independent of C_j. Then there are about 1/10 and $\frac{1}{3}$ censored observations. The hazard rate is a strictly decreasing function on $[0, 1]$ with range $[0, +\infty)$.

Figure 2 shows that the performance of the LP estimator (left panel) and M&W's (right panel) in the 400 simulations. The LP estimator suc-

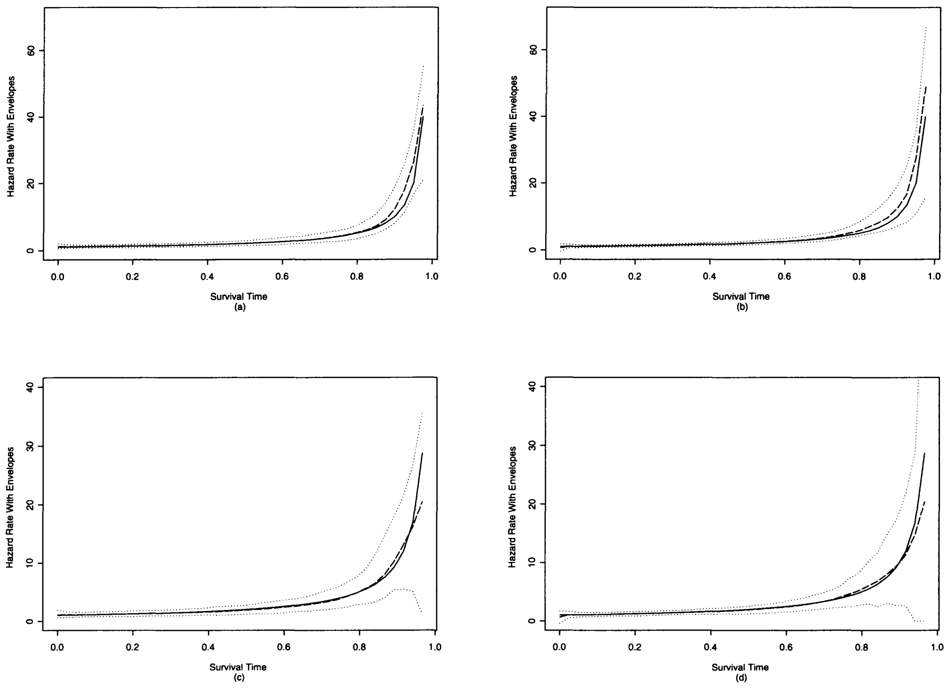

Figure 1. Simulation results for Example 5.1. (a)–(b): 10% censoring; (c)–(d): 30% censoring. Left panels—LP estimator; right panels—M&W's estimator. Solid line—true hazard rate, dashed line—pointwise median curve among 400 simulations, dotted lines—envelopes formed via pointwise 2.5% and 97.5% sample percentiles.

ceeds better in capturing the structure of the real hazard rate near the left boundary, in term of the 5% empirical confidence bands.

Example 5.3. Note that bathtub hazard rate is often encountered in practice in reliability. See Nelson (1990) and Höyland and Rausland (1994). We simulated 400 samples of size $n = 250$ from the bathtub-shaped model, the "quadratic concave up" hazard considered in Hess *et al.* (1999):

$$\lambda(t) = 0.1277\left(\frac{t^2}{2500} - \frac{t}{25} + 1\right), \quad t \in [0, 100].$$

The constant 0.1277 was chosen to leave about 10 units at risk at $t = 90$, i.e., $P(T \geq 90) = 0.04$. Figure 3 presents the median performance of the LP estimator and M&W's among the 400 simulations, together with their envelopes formed by pointwise 2.5% and 97.5% sample percentiles under $1/10$ and $\frac{1}{3}$ censoring based on the Lehmann model. The performance of the LP estimator is satisfactory.

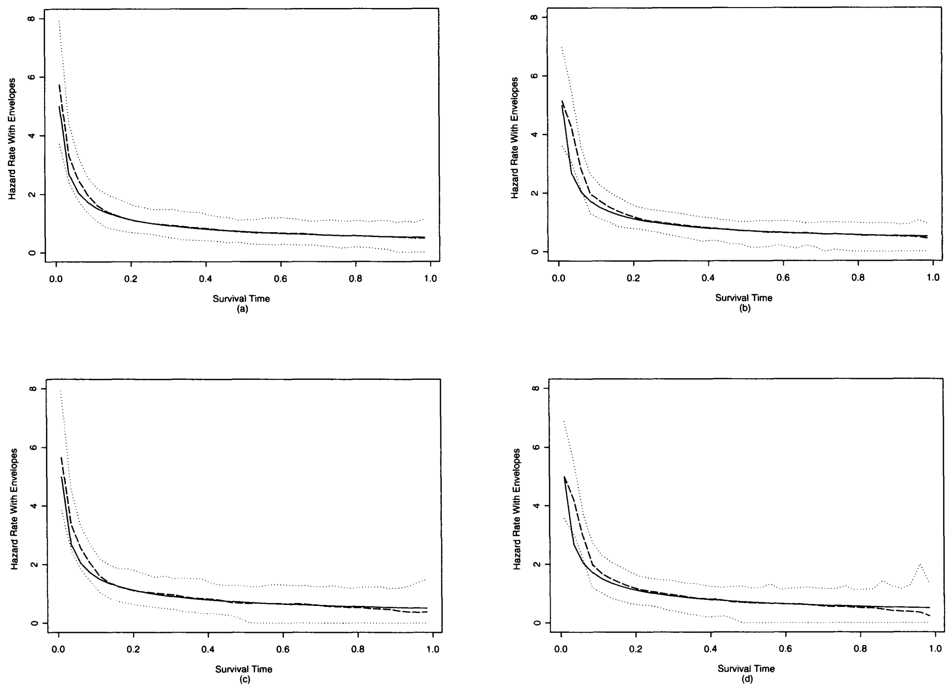

Figure 2. Simulation results for Example 5.2. (a)–(b): 10% censoring; (c)–(d): 30% censoring. Left panels—LP estimator; right panels—M&W's estimator. Solid line—true hazard rate, dashed line—pointwise median curve among 400 simulations, dotted lines—envelopes formed via pointwise 2.5% and 97.5% sample percentiles.

Table 1. Average of the estimated squared bias, variance and MSE under 10% censoring

Ex1	Bias^2	Variance	MSE
LP	1.80	2.62	4.42
M&W	3.34	4.76	8.10
Ex2	Bias^2	Variance	MSE
LP	0.032	0.074	0.106
M&W	0.083	0.063	0.146
Ex3	Bias^2	Variance	MSE
LP	1.22e-006	6.53e-005	6.65e-005
M&W	2.26e-006	7.550e-005	7.78e-005

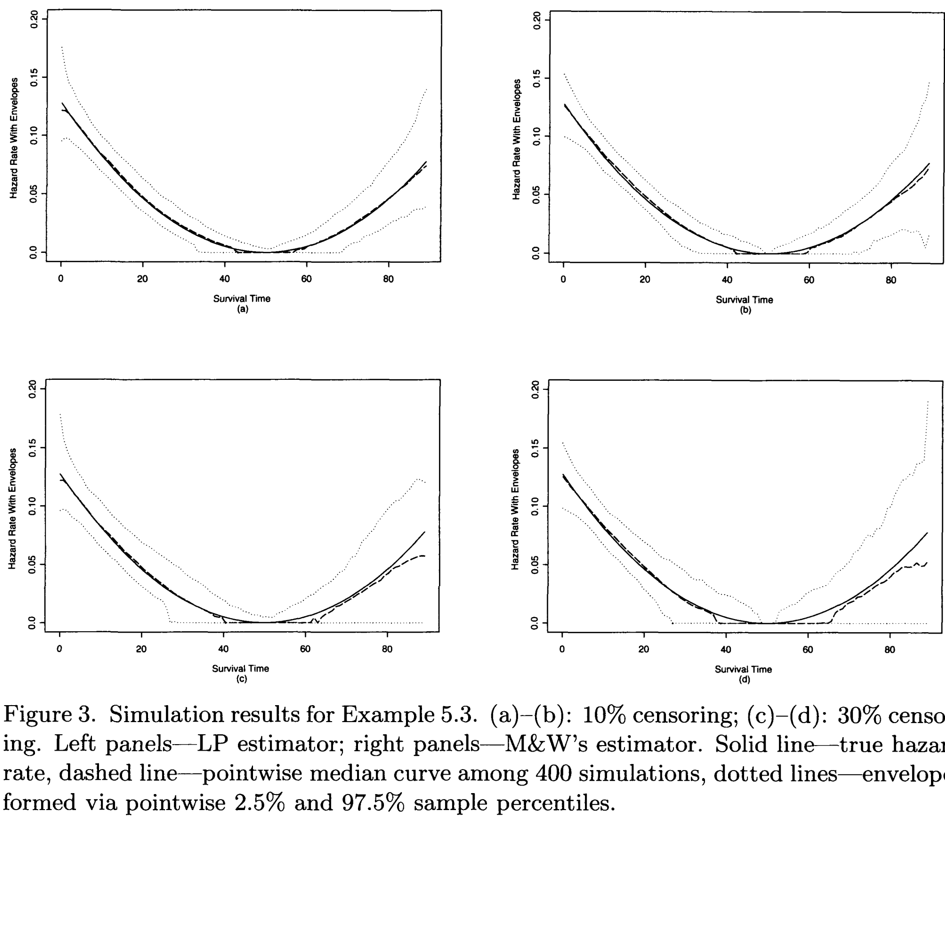

Figure 3. Simulation results for Example 5.3. (a)–(b): 10% censoring; (c)–(d): 30% censoring. Left panels—LP estimator; right panels—M&W's estimator. Solid line—true hazard rate, dashed line—pointwise median curve among 400 simulations, dotted lines—envelopes formed via pointwise 2.5% and 97.5% sample percentiles.

Table 2. Average of the estimated squared bias, variance and MSE under $\frac{1}{3}$ censoring

Ex1	Bias^2	Variance	MSE
LP	2.15	3.41	5.56
M&W	1.51	7.53	9.04
Ex2	Bias^2	Variance	MSE
LP	0.035	0.104	0.139
M&W	0.090	0.100	0.190
Ex3	Bias^2	Variance	MSE
LP	2.03e-005	1.78e-004	1.98e-004
M&W	2.90e-005	2.25e-004	2.54e-004

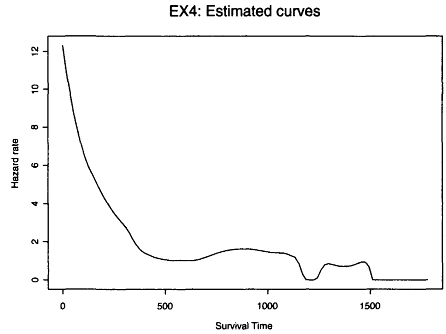

Figure 4. The LP estimator of the hazard rate for the Stanford heart transplant data.

To understand more about the performance of LP and M&W's estimators, we computed pointwise bias, variances and MSE's for the two estimators over certain grid points. Tables 1 and 2 report the average of the estimated squared bias, variances and MSE's for the LP and M&W's estimators over the grid points we considered for the above three examples, under 1/10 and $\frac{1}{3}$ censoring, respectively. It shows that for these models the LP estimator compares favorably in terms of its MSE's to M&W's estimators.

Example 5.4. In this example, we use the Stanford heart transplant data from Kalbfleisch and Prentice (1980). We estimate the hazard function by the data-driven method mentioned in our algorithm. Figure 4 gives the estimated curve of hazard rate function. Our nonparametrically estimated curve clearly suggests a Weibull model to be appropriate for the data, which is consistent with most parametric analyses of the data, say, in Kalbfleisch and Prentice (1980).

Acknowledgements. The work of J. Jiang is supported by the Chinese NSF grants 10001004 and 39930160. The work of K. Doksum is supported in part by NSF grant DMS-9971301.

APPENDIX

Proofs of Theorems 3.1 *and* 3.2. (i) Note that by (2.1) and (2.8)

$$\text{(A.1)} \quad S_{n\ell}(x_0) = \int \frac{1}{b_n} K\left(\frac{u - x_0}{b_n}\right)\left(\frac{u - x_0}{b_n}\right)^{\ell} d\Lambda_n(u)$$

$$
\begin{aligned}
&= \int \frac{1}{b_n} K\Big(\frac{u-x_0}{b_n}\Big)\Big(\frac{u-x_0}{b_n}\Big)^{\ell} d\Lambda(u) \\
&\qquad + \int \frac{1}{b_n} K\Big(\frac{u-x_0}{b_n}\Big)\Big(\frac{u-x_0}{b_n}\Big)^{\ell} d[\Lambda_n(u) - \Lambda(u)] \\
&\equiv \beta_{n\ell}(x_0) + \gamma_{n\ell}.
\end{aligned}
$$

Let $\beta_{\mathbf{n}}(x_0) = (\beta_{n0}, \ldots, \beta_{np})^T$ and $\gamma_{\mathbf{n}}(x_0) = (\gamma_{n0}, \ldots, \gamma_{np})^T$. Then

$$
\text{(A.2)} \qquad \mathbf{S_n}(x_0) = \beta_{\mathbf{n}}(x_0) + \gamma_{\mathbf{n}}(x_0).
$$

From (2.7), we know

$$
\begin{aligned}
\text{(A.3)} \qquad \mathbf{B}[\hat{\mathbf{a}}(x_0) - \mathbf{a}(x_0)] &= \mathbf{S}^{-1}\boldsymbol{\beta}_n(x_0) - \mathbf{B}\mathbf{a}(x_0) + \mathbf{S}^{-1}\boldsymbol{\gamma}_n(x_0) \\
&\equiv \boldsymbol{\alpha}_n(x_0) + \mathbf{S}^{-1}\boldsymbol{\gamma}_n(x_0).
\end{aligned}
$$

We will show that $\boldsymbol{\alpha}_n(x_0)$ contributes to the bias term of the LP estimator, and $\mathbf{S}^{-1}\boldsymbol{\gamma}_n(x_0)$ to the variance term.
First, by (A.1), change of variable for integration, and a Taylor expansion, we get, for $\ell = 0, 1, \ldots, p$,

$$
\begin{aligned}
\beta_{n\ell}(x_0) &= \int K(t) t^{\ell} \lambda(x_0 + b_n t)\, dt \\
&= \sum_{j=0}^{p} \frac{b_n^j}{j!} \lambda^{(j)}(x_0) s_{\ell+j} + \frac{b_n^{p+1}}{(p+1)!} \lambda^{(p+1)}(x_0) s_{\ell+p+1} \\
&\qquad + \frac{b_n^{p+2}}{(p+2)!} \lambda^{(p+2)}(x_0) s_{\ell+p+2} + o(b_n^{p+2}),
\end{aligned}
$$

Then

$$
\begin{aligned}
\text{(A.4)} \quad \boldsymbol{\alpha}_n(x_0) &= \frac{b_n^{p+1}}{(p+1)!} \lambda^{(p+1)}(x_0) \mathbf{S}^{-1} \mathbf{c}_p \\
&\qquad + \frac{b_n^{p+2}}{(p+2)!} \lambda^{(p+2)}(x_0) \mathbf{S}^{-1} \mathbf{c}_{p+1} + o(b_n^{p+2}),
\end{aligned}
$$

in particular, if $\lambda(x)$ is a polynomial up to order p in a neighbourhood of x_0, then the exact bias $\boldsymbol{\alpha}_n(x_0)$ of the LP estimator is zero.
Second, we know by Lo et al. (1989)

$$
\text{(A.5)} \qquad \Lambda_n(x) - \Lambda(x) = \frac{1}{n} \sum_{i=1}^{n} \xi(X_i, \delta_i, x) + r_n(x),
$$

where for $z \geq 0$, $x \geq 0$, and $\delta = 1$ or 0,

$$
\sup_{0 \leq x \leq T} |r_n(x)| = O\Big(\frac{\log n}{n}\Big), \quad \text{a.s.}
$$

$$\xi(z,\delta,x) = g\big(\min(z,x)\big) - I(z \le x, \delta = 1)/\bar{L}(x),$$

and

$$g(x) = \int_0^x [\bar{L}(u)]^{-2}\, dL_1(u).$$

Note that

$$E\xi(X_i,\delta_i,x) = 0, \tag{A.6}$$

and

$$\operatorname{Cov}\big(\xi(X_i,\delta_i,s), \xi(X_i,\delta_i,t)\big) = g\big(\min(s,t)\big). \tag{A.7}$$

Let $K_\ell(t) = K(t)t^\ell$. Using the definition of $\gamma_{n\ell}$, (A.5) and integration by parts, we obtain the following almost surely representation of $\gamma_{n\ell}(x_0)$:

$$\gamma_{n\ell}(x_0) = \sigma_{n\ell}(x_0) + e_{n\ell}(x_0), \tag{A.8}$$

where

$$\begin{aligned}\sigma_{n\ell}(x_0) &= \frac{1}{n}\sum_{i=1}^n \int \frac{1}{b_n} K\Big(\frac{u-x_0}{b_n}\Big)\Big(\frac{u-x_0}{b_n}\Big)^\ell d\xi(X_i,\delta_i,u)\\ &= (nb_n)^{-1}\sum_{i=1}^n \int \xi(X_i,\delta_i,x_0+b_nt)dK_\ell(t)\end{aligned}$$

is the stochastic component of $S_{n\ell}(x_0)$ and contributes to the variance of the LP estimator, and $e_{n\ell}(x_0)$ is the negligible error of the approximation which satisfies

$$\sup_{0\le x_0 \le T} |e_{n\ell}(x_0)| = O(\frac{\log n}{n}), \quad \text{a.s.} \tag{A.9}$$

for $0 \le \ell \le p$. Note that $E(\sigma_{n\ell}(x_0)) = 0$ and

$$\begin{aligned}&\operatorname{Cov}(\sigma_{n\ell}(x_0), \sigma_{nm}(x_0)) \\ &= \frac{1}{nb_n}\int\int \frac{1}{b_n} g(\min(x_0+b_nu, x_0+b_nv))\, dK_\ell(u)\, dK_m(v)\\ &= \frac{1}{nb_n^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{x_0+b_nv} K_\ell\Big(\frac{u-x_0}{b_n}\Big)\, dg(u)\, dK_m(v)\\ &= \frac{1}{nb_n^2}\int_{-\infty}^{+\infty} K_\ell\Big(\frac{u-x_0}{b_n}\Big) K_m\Big(\frac{u-x_0}{b_n}\Big)\, dg(u)\\ &= \frac{1}{nb_n}\int_{-\infty}^{+\infty} \frac{1}{b_n}K^2\Big(\frac{u-x_0}{b_n}\Big)\Big(\frac{u-x_0}{b_n}\Big)^{\ell+m} dg(u)\\ &= \frac{1}{nb_n}\int K^2(t)t^{\ell+m}\frac{\lambda(x_0+b_nt)}{\bar{L}(x_0+b_nt)}\, dt\\ &= \frac{1}{nb_n}\frac{\lambda(x_0)}{\bar{L}(x_0)} v_{\ell+m}(1+o(1)).\end{aligned} \tag{A.10}$$

Let $\boldsymbol{\sigma}_n(x_0) = (\sigma_{n0}, \ldots, \sigma_{np})^T$ and $\mathbf{e_n}(x_0) = (e_{n0}, \cdots, e_{np})^T$. Then by (A.8) and (A.10)

$$\sqrt{nb_n}\mathbf{S}^{-1}\boldsymbol{\gamma}_n(x_0) = \sqrt{nb_n}\mathbf{S}^{-1}\boldsymbol{\sigma}_n(x_0) + \sqrt{nb_n}\mathbf{S}^{-1}\mathbf{e}_n(x_0) \tag{A.11}$$

and

$$\mathrm{Cov}\big(\boldsymbol{\sigma}_n(x_0), \boldsymbol{\sigma}_n(x_0)\big) = \frac{1}{nb_n}\frac{\lambda(x_0)}{\bar{L}(x_0)}\mathbf{V}^*\big(1 + o(1)\big). \tag{A.12}$$

By the central limit theorem, we get

$$\sqrt{nb_n}\boldsymbol{\sigma}_n(x_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{\lambda(x_0)}{\bar{L}(x_0)}\mathbf{V}^*). \tag{A.13}$$

By (A.9), we know

$$\sup_{0 \le x_0 \le T} |\mathbf{S}^{-1}\mathbf{e_n}(x_0)| = O\Big(\frac{\log n}{n}\Big), \quad \text{a.s.} \tag{A.14}$$

Then by (A.12), (A.13) and (A.14)

$$\sqrt{nb_n}\mathbf{S}^{-1}\boldsymbol{\gamma}_n(x_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{\lambda(x_0)}{\bar{L}(x_0)}\mathbf{S}^{-1}\mathbf{V}^*\mathbf{S}^{-1}). \tag{A.15}$$

Combination of (A.3), (A.4) and (A.15) completes the proof of Theorem 3.2.

(ii) Note that $\sigma_{n\ell}(x_0)$, for $\ell = 0, \ldots, p$, are i.i.d.'s sums. By the SLN, we know $\sigma_{n\ell}(x_0) \xrightarrow{\text{a.s.}} E(\sigma_{n\ell}(x_0)) = 0$. Then $\boldsymbol{\sigma}_n(x_0) \xrightarrow{\text{a.s.}} 0$. This combined with (A.8) and (A.9) yields $\boldsymbol{\gamma}_n(x_0) \xrightarrow{\text{a.s.}} 0$. Therefore, by (A.3) and (A.4),

$$\mathbf{B}[\hat{\mathbf{a}}(x_0) - \mathbf{a}(x_0)] \xrightarrow{\text{a.s.}} 0. \qquad \square$$

Proof of Theorem 3.3. Note that (a) $s_j = 0$ for j odd, (b) $(\mathbf{S})_{ij} = (\mathbf{S}^{-1})_{ij}$ for $i + j$ even. Simple algebra gives that $\mathbf{e}_k^T\mathbf{S}^{-1}\mathbf{c}_p = 0$ for $p - k$ even, then the theorem follows from Theorem 3.2. $\square$

Proof of Remark 3.2. Remark 3.2 follows by the same argument as in (i). $\square$

REFERENCES

Aarset, M.V. (1987). How to identify a Bathtub hazard rate, *IEEE Transactions on Reliability* R-36, 106–108.

Efron, B. (1988). Logistic regression, survival analysis, and the Kaplan-Meier curve, *J. Amer. Statist. Assoc.* 83, 414–425.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption, *J. Roy. Statist. Soc. B* 57, 371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions, In *Smoothing Techniques for Curve Estimation*, pp. 23–68. Volume 757 of Lecture Notes in Math. Springer, New York.

González-Manteiga, W., Cao, R., and Marron, J.S. (1996), Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation, *J. Amer. Statist. Assoc.* 91, 1130–1140.

Gross, S.T. and Lai, T.L. (1996). Nonparametric estimation and regression analysis with left-truncated and right censored data, *J. Amer. Statist. Assoc.*, 91, 1166–1180.

Hall, P. and Wehrly, T.E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators, *J. Amer. Statist. Assoc.* 86, 665–672.

Hess, K.R., Serachitopol, D.M., and Brown, B.W. (1999). Hazard function estimators: a simulation study, *Statistics in Medicine* 18, 3075–3088.

Höyland, A. and Rausland, M. (1994). *System Reliability: Models and Statistical Methods*, Wiley, New York.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation, *Statistics and Computing* 3, 135–146.

Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time data*, Wiley, New York.

Koziol, J. and Green, S. (1976), A Cramér-von Mises statistic for randomly censored data, *Biometrika* 63, 465–474.

Lejeune, M. and Sarda, P. (1992), Smooth estimators of distribution and density functions, *Computational Statistics and Data Analysis* 14, 457–471.

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, *Prob. Th. Rel. Fields* 80, 461–473.

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error, *Ann. Statist.* 20, 712–736.

Mudholkar, G.S., Srivastava, D.K., and Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data, *J. Amer. Statist. Assoc.* 91, 1575–1583.

Müller, H.-G. and Wang, J.L. (1990). Locally adaptive hazard smoothing, *Prob. Th. Rel. Fields*, 85, 523–538.

Müller, H.-G. and Wang, J.L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths, *Biometrics* 50, 61–76.

Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analysis.* Wiley, New York.

Nielsen, J.P. and Tanggaard, C. (2001), Boundary and bias correction in kernel hazard estimation, *Scand. J. Statist.* 28, 675–698.

Patil, P.N. (1993). On the least squares cross-validation bandwidth in hazard rate estimation, *Ann. Statist.* 21, 1792–1810.

Ramlau-Hausen, H. (1983). Smoothing counting process intensities by means of kernel functions, *Ann. Statist.* 11, 453–466.

Tanner, M. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method, *Ann. Statist.* 11, 989–993.

Yandell, B.S. (1983). Nonparametric inference for rates with censored survival data, *Ann. Statist.* 11, 1119–1135.

Jiancheng Jiang
Dept. of Probability & Statistics
Peking University
Beijing 100871
P.R. China
jiang@math.pku.edu.cn

Kjell Doksum
Department of Statistics
University of California
Berkeley, CA 94720-3860
USA
doksum@stat.Berkeley.edu