

On local polynomial estimation of hazard rates and their derivatives under left truncation and right censoring models

Abstract

Estimating hazard rate function is an important problem in survival analysis. There are some estimation approaches based on kernel smoothing. However, they suffer from the boundary effects or need high order kernels, which increases the mean squared error. We introduce local polynomial estimators of hazard rates and their derivatives for the left truncation and right censoring models. The estimators have favorable properties similar to those of local polynomial regression estimators. Asymptotic expressions for the mean squared errors (AMSE's) are obtained. Consistency and joint asymptotic normality of the local polynomial estimators are established. A data-based local bandwidth selection rule is proposed.

Keywords: censoring, data-driven local bandwidth selection, hazard rate, local polynomial, truncation

Volume 7 Issue 3 - 2018

Jiancheng Jiang,¹ Lingju Chen,² Yuze Yuan²

¹Department of Mathematics and Statistics, University of North Carolina at Charlotte, USA

²Department of Mathematics, Minjiang University, China

Correspondence: Yuze Yuan, Department of Mathematics, Minjiang University, China. Email yuanyuze@126.com

Received: April 28, 2018 | **Published:** May 15, 2018

Introduction

Consider a subject in survival studies. Only if its onset time, i.e. the time origin of its lifetime, passes the beginning of the study, the subject can enter into the study. For the entered individuals, each of them is then followed for a fixed time point. Such subjects are so-called left truncated and right censored. To be specific, let X be the lifetime with distribution function (df) $F(x)$, T the random left truncation time with arbitrary df $G(x)$, and C the random censoring time with arbitrary df $L(x)$, where x is independent of (T, C) . Then the cumulative hazard function of $F(x)$ is $\Lambda(x) = \int_0^x dF(t) / [1 - F(t)]$. Under the left truncation and right censoring model, one observes (Y, T, δ) if $Y \geq T$, where $Y = \min(X, C)$ and $\delta = I(X \leq C)$ is an indicator of the censoring status of X which takes value one if $X \leq C$ and zero otherwise. When $Y < T$, nothing is observed (see for example, Gurler & Wang¹). Let the distribution of Y be $W(y)$ and assume that $\alpha = P(T \leq Y) > 0$. Then $\bar{W} = \bar{F}L$, where and throughout the paper for any df. E , $\bar{E} = 1 - E$ is the corresponding survival function.

The left truncation and right censoring model has been investigated by many authors. Interesting work along the field can be found in Gross & Lai² and Gurler & Wang¹ among others. Several authors have considered the estimation of hazard functions under the left truncation and right censoring model. For examples, Uzunogullari & Wang,³ and Wu & Wells.⁴

In the present investigation, we study local polynomial (LP for short) estimators of hazard functions and their derivatives based on the i.i.d sample $\{(Y_i, T_i, \delta_i)\}_{i=1}^n$ from (Y, T, δ) . Under the left truncation and right censoring model, one observes only those i.i.d pairs (Y_i, T_i, δ_i) for which $Y_i \geq T_i$. It is worth pointing out that our estimators inherit some favorable properties from local polynomial regression estimators, in particular, our estimators can reduce the bias according to the degree of the polynomial without increasing the variance and automatically correct the left boundary effect. The point wise asymptotic normality of our estimators enables one to find

the asymptotically optimal variable bandwidth choice, and thereafter allows one to develop a data-driven optimal local bandwidth selector by using the ideas of Fan & Gijbels.⁵ We here present a simpler data-driven method for choosing the local bandwidth.

The outline of this paper is as follows. In Section 2, we introduce the LP estimators. In Section 3 we concentrate on the asymptotic properties of the proposed estimators, including point wise strong consistency and joint asymptotic normality. In Section 4, we propose the data-driven local bandwidth selection rule. Technical proofs are given in the Appendix.

Estimation

In order to introduce the LP estimators, we use the following notation from Gurler & Wang¹

$$(i) \quad C(z) = P(T \leq z \leq Y | Y \geq T) = \alpha^{-1} P(T \leq z \leq C) [1 - F(z)]$$

$$(ii) \quad C_n(z) = (n+1)^{-1} \sum_{i=1}^n I(T_i \leq z \leq Y_i), \text{ the modified empirical estimator of } C(z).$$

$$(iii) \quad F_n(x) = 1 - \prod_{Y_i \leq x} (1 - [nC_n(Y_i)]^{-1})^{\delta_i}, \text{ the product-limit estimator of } F(x) \text{ from Tsai, et al.}^{18}$$

Let $a_E = \inf\{t : E(t) > 0\}$ and $b_E = \inf\{t : E(t) = 1\}$ denote the left and right endpoints of the support for any d.f. $E(x)$, respectively. Then $F(x)$ is identifiable if $a_G \leq a_W$ and $b_G \leq b_W$ (Woodrooffe⁶ and Gurler & Wang¹).

Therefore, we assume this condition holds. As in Gurler & Wang,¹ for estimating the density function $f(x)$ of X , we also assume $a_G \leq a_W$.

Following Gurler & Wang,¹ we define $W_1(y) = P(Y \leq y, \delta = 1 | Y \geq T)$ and $W_{1n}(y) = n^{-1} \sum_{i=1}^n I(Y_i \leq y, \delta_i = 1)$. Then $dW_1(y) = \alpha^{-1} P(T \leq Y \leq C) dF(y)$ and

$$\Lambda(x) = \int_0^x dW_1(y) / C(y).$$

By Gurler & Wang,¹ the Nelson-Aalen type estimator of $\Lambda(x)$ is given by

$$\Lambda_n(x) = \int_0^x [C_n(y)]^{-1} dW_{1n}(y) = \sum_{i=1}^n \frac{I(Y_i \leq x, \delta_i = 1)}{nC_n(Y_i)}. \tag{1}$$

Gurler & Wang¹ considered the following kernel estimator for the density function $f(x)$ of X , which is a convolution of the product-limit estimator $F_n(x)$ with an appropriate kernel function K_ν :

$$\hat{f}^{(\nu)}(x) = \frac{1}{b_n^{\nu+1}} [K_\nu(\frac{x-u}{b_n}) dF_n(u), \tag{2}$$

where K_ν is a higher order kernel. The method can be also adapted to the case of estimating the hazard function $\lambda(x)$ and its derivatives if one uses $\Lambda_n(x)$ instead of $F_n(x)$:

$$\hat{\lambda}(x) = \frac{1}{b_n^{\nu+1}} [K_\nu(\frac{x-u}{b_n}) d\Lambda_n(u). \tag{3}$$

The estimator is an extension to that of Müller & Wang^{7,8} where right censoring model is considered. However, 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 & Wand.⁹ When estimating at point x near a_G or b_W , the effective support $[x - b_n, x + b_n]$ of the kernel is not contained in $[a_G, b_W]$, most kernel estimators in density estimation and regression settings will encounter boundary effects. The estimator (3) suffers from boundary effects near the endpoints of the support of the hazard rates. In the presence of censoring for estimating hazard rate function, Müller & Wang⁸ solved the problem by employing boundary kernels and a data-adaptive varying bandwidth selection procedure. Hall & Wehrly¹⁰ used a geometrical method for removing edge effects from kernel-type nonparametric regression estimators. These boundary correction methods may also be adapted to the estimation approach in (3). Here we introduce a simple and intuitive approach to the problem. Our approach does not need higher order kernels or boundary kernels while automatically correcting the boundary effects. Our idea is similar to that of Jiang & Doksum,¹¹ but their procedure cannot be directly applied to the current setting.

Following Jiang and Doksum,¹¹ we consider the following optimization problem:

$$\min_{a_j} \int \frac{1}{b_n} K(\frac{u-x}{b_n}) [\lambda(u) - \sum_{j=0}^p a_j (u-x)^j]^2 du. \tag{4}$$

By Taylor expansion, the solution of the optimization problem, $\mathbf{a}^*(x) \equiv (a_0^*, \dots, a_p^*)^T$, will estimate $\mathbf{a}(x) \equiv (\lambda(x), \dots, \lambda^{(p)}(x) / p!)^T$. Since $\Lambda_n(x)$ in (1) is the empirical estimator of $\Lambda(x)$, we define the following *generalized empirical hazard rate* as the generalized derivative of $\Lambda_n(x)$:

$$\lambda_n(x) = \sum_{i=1}^n \frac{D(x - Y_i) \delta_i}{nC_n(Y_i)}, \tag{5}$$

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

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

for any integral function $g(x)$. Then $\int_0^x \lambda_n(t) dt = \Lambda_n(x)$, which is why we call λ_n the *generalized empirical hazard rate*. Replacing $\lambda(x)$ in (4) by $\lambda_n(x)$, we obtain that

$$\hat{\mathbf{a}}(x) \equiv (\hat{a}_0, \dots, \hat{a}_p)^T = \arg \min_{a_j} \int \frac{1}{b_n} K(\frac{u-x}{b_n}) [\lambda_n(u) - \sum_{j=0}^p a_j (u-x)^j]^2 du. \tag{6}$$

Then $\hat{\mathbf{a}}(x)$ is the *LP estimator* of $\mathbf{a}(x)$. Jones¹² considered a locally linear estimator and established its link with the generalized jackknife boundary correction for $p=1$ by using the ideas of Lejeune & Sarda¹³ in local linear fitting to the empirical distribution F_n . Here we study the local polynomial estimation of hazard functions and their derivatives, $\mathbf{a}(x)$, under the left truncation and right censoring model. Obviously, our method can be used for the complete data case, which corresponds to “no truncation and no censoring”.

Taking the derivative with respect to a 's of the integral in (6), we obtain the *LP estimator* $\hat{\mathbf{a}}(x)$ as the solution to the linear equations: for $\ell = 0, \dots, p$,

$$\sum_{i=1}^n \frac{1}{b_n} K(\frac{Y_i-x}{b_n}) (Y_i-x)^\ell \frac{\delta_i}{nC_n(Y_i)} = \sum_{i=0}^p a_i \int_{u \geq 0} (u-x)^{i+\ell} \frac{1}{b_n} K(\frac{u-x}{b_n}) du. \tag{7}$$

It follows that the *LP estimator* at any point $x_0 \in (a_G, b_W)$ has the following closed form:

$$\mathbf{B}\hat{\mathbf{a}}(x_0) = \mathbf{S}^{-1} \mathbf{S}_n(x_0), \tag{8}$$

where $\mathbf{S}_n(x_0) = (S_{n0}(x_0), \dots, S_{np}(x_0))^T$, and

$$S_{n\ell}(x_0) = \sum_{i=1}^n \frac{1}{b_n} K(\frac{Y_i-x_0}{b_n}) (\frac{Y_i-x_0}{b_n})^\ell \frac{\delta_i}{nC_n(Y_i)}. \tag{9}$$

When $p=0$ and $s_0=1$,

$$\hat{a}_0 = \sum_{i=1}^n \frac{1}{b_n} K(\frac{Y_i-x_0}{b_n}) \frac{\delta_i}{nC_n(Y_i)} = \frac{1}{b_n} [K(\frac{x_0-u}{b_n}) d\Lambda_n(u),$$

which is the same as the kernel estimator of $\lambda(x_0)$ in (3). However, this equivalence does not hold for boundary points.

We will show in next section that the *LP estimator* shares nice properties with the local polynomial regression estimator, in particular, the estimator will keep its convergence rate up to the left boundary point, i.e. the estimator automatically corrects the left boundary effect, which contrasts with the results for other hazard rate estimators.

Asymptotic properties

In this section, we will establish the consistency and joint asymptotic normality of the local polynomial estimators. To this end, we introduce some regularity conditions. For a given point

$x_0 \in (a_G, b_W)$, the following notations and assumptions are needed.

- (A1) The hazard rate function $\lambda(x) = \Lambda'(x)$ has a continuous $(p+1)$ th derivative at the point x_0 .
- (A2) The sequence of bandwidths b_n tends to zero such that $nb_n \rightarrow +\infty$ as $n \rightarrow +\infty$. Let $\mathbf{B} = \text{diag}(1, b_n, \dots, b_n^p)$.
- (A3) $C(x)$ is continuous at the point x_0 .
- (A4) The kernel function $K(x)$ is a continuous function of bounded variation and with bounded support $[-1, 1]$, say. Let $s_\ell = \int_{-1}^1 K(u)u^\ell du$, $v_\ell = \int_{-1}^1 u^\ell K^2(u)du$, for $\ell \geq 0$, $\mathbf{c}_p = (s_{p+1}, \dots, s_{2p+1})^T$, $\mathbf{S} = (s_{i+j})$ and $\mathbf{V}^* = (v_{i+j})$ (for $0 \leq i \leq p$; $0 \leq j \leq p$) be $(p+1) \times (p+1)$ matrices.

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

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

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

$$\begin{aligned} & \sqrt{nb_n} [\mathbf{B}(\hat{\mathbf{a}}(x_0) - \mathbf{a}(x_0)) - \frac{\lambda^{(p+1)}(x_0)b_n^{p+1}}{(p+1)!} \mathbf{S}^{-1} \mathbf{c}_p (1 + o_p(1))] \\ & \xrightarrow{L} N(0, \mathbf{S}^{-1} \mathbf{V}^* \mathbf{S}^{-1} \lambda(x_0) C^{-1}(x_0)). \end{aligned} \tag{10}$$

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^{7,8} based on higher order kernels, for which the respective zero bias only holds true asymptotically.

Remark 3.2

When there is no truncation, we take $T = 0$ and $G(x) = 1$ over the support of $F(x)$, then $C(x) = \alpha^{-1} P(C \geq x) \bar{F}(x) = \bar{W}(x)$, and the asymptotic normality for interior point x_0 is the same as in Müller & Wang.^{7,8} When there is no censoring, we take the censoring variable $C = +\infty$ and $L(x) = 0$ over the support of $F(x)$, then $C(x) = P(T \leq x \leq X | Y \geq T)$, and the asymptotic normality for interior point x_0 is the same as in Gürlér & Wang.¹

From Theorem 3.2, the asymptotic mean squared error AMSE for estimating $k!$ $\lambda^{(k)}(x_0)$ ($k = 0, \dots, p$) is

$$AMSE_k(b_n, x_0) = b_n^{2(p+1-k)} \left[\mathbf{e}_k^T \mathbf{S}^{-1} \mathbf{c}_p \frac{\lambda^{(p+1)}(x_0)}{(p+1)!} \right]^2 + \frac{1}{nb_n^{2k+1}} \mathbf{e}_k^T \mathbf{S}^{-1} \mathbf{V}^* \mathbf{S}^{-1} \mathbf{e}_k \frac{\lambda(x_0)}{C(x_0)} \tag{11}$$

Where $\mathbf{e}_k = (0, \dots, 1, \dots, 0)^T$ has one in the $k+1$ th component and zeros in the others. Therefore, the optimal local bandwidth for estimating the k th derivative of $\lambda(x)$ at x_0 , in the sense of minimizing $AMSE_k(b_n, x_0)$, is

$$b_{k,opt}(x_0) = n^{-\frac{1}{2p+3}} \left(\frac{[(p+1)!]^2 \mathbf{e}_k^T \mathbf{S}^{-1} \mathbf{V}^* \mathbf{S}^{-1} \mathbf{e}_k \lambda(x_0) / C(x_0)}{2(p+1-k) [\lambda^{(p+1)}(x_0)]^2 (\mathbf{e}_k^T \mathbf{S}^{-1} \mathbf{c}_p)^2} \right)^{1/(2p+3)} \tag{12}$$

Theorem 3.3

Consider the left edge effect on the estimator. 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 (8), $\hat{\mathbf{a}}(x_n)$ in (6) has the following closed form, for $p = 0, 1, 2, \dots$,

$$\mathbf{B}\hat{\mathbf{a}}(x_n) = \mathbf{S}_d^{-1} \mathbf{S}_n(x_n), \tag{13}$$

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 (10) continues to hold with \mathbf{c}_p , \mathbf{S} and \mathbf{V}^* replaced by $\mathbf{c}_{p,d}$, \mathbf{S}_d and \mathbf{V}_{d}^* , respectively, where $\mathbf{c}_{p,d} = (s_{p+1,d}, \dots, s_{2p+1,d})^T$ and $\mathbf{V}_{d}^* = (v_{i+j,d})$ is $(p+1) \times (p+1)$ matrices.

This property of our estimator in Theorem 3.3 is similar to that of local polynomial regression estimation, which is not shared by other kernel estimators of hazard rates (Hess et al.¹⁴) The LP estimators are automatically boundary adaptive in the sense of Fan & Gijbels.¹⁵ Note that the above property holds even for $p = 0$, which contrasts with the cases of local polynomial regression.

Remark 3.3 For a finite sample, one may encounter right boundary effects when estimating near T . A good method for dealing with the problem is to use the following estimator similar to the estimator in (13):

$$\mathbf{B}\hat{\mathbf{a}}(x_0) = \mathbf{S}_q^{-1} \mathbf{S}_n(x_0), \tag{14}$$

where $x_0 = T - qb_n$ for $q \in [0, 1]$, and \mathbf{S}_q defined as \mathbf{S} but with s_i replaced by $s_{i,q} = \int_{-q}^1 u^i K(u) du$.

Data-driven local bandwidth choice

The proposed estimators depend on the band width b_n . It is important to develop a local bandwidth choice for estimating hazard functions and their derivatives, especially when one would like to have a data-driven approach to bandwidth choice in practice. For hazard rate estimation, Patil¹⁶ considered least squares cross-validation bandwidth selection, and Gonzalez-Manteiga et al.¹⁷ studied smoothed bootstrap selection of the global bandwidth. Müller & Wang,⁷ Hess et al.¹⁴ and Jiang & Doksum¹¹ studied the local bandwidth choice for estimating hazard rates under right censoring. Here we extend the data-driven local bandwidth choice of Jiang & Doksum¹¹ to the left truncation and right censoring model.

From the proof of Theorem 3.2, we see that the exact bias of the estimator $\hat{a}_k(x_0)$ is

$$B_k(b_n, x_0) = \mathbf{e}_k^T (\mathbf{S}^{-1} \beta_n(x_0) - \mathbf{B}\mathbf{a}(x_0)), \tag{15}$$

where $\beta_n(x_0) = (\beta_{n0}(x_0), \dots, \beta_{np}(x_0))^T$ and $\beta_{nk}(x_0) = \int K(t)^k \lambda(x_0 + b_n t) dt$. The asymptotic variance of $\hat{a}_k(x_0)$ is

$$V_k(b_n, x_0) = \frac{1}{nb_n} \mathbf{e}_k^T \mathbf{S}^{-1} \tilde{\mathbf{V}} \mathbf{S}^{-1} \mathbf{e}_k, \tag{16}$$

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

$$AMSE_k(b_n, x_0) = \hat{B}_k^2(b_n, x_0) + \hat{V}_k(b_n, x_0), \tag{17}$$

where $\hat{B}_k(b_n, x_0)$ and $\hat{V}_k(b_n, x_0)$ 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 $C(x)$ replaced by its empirical estimator $C_n(x)$. We define

$$\hat{b}_{k,opt}(x_0) = \arg \min_b AMSE_k(b, x_0).$$

Then, for estimating $\lambda(x)$ when $p=1$, the following algorithm similar to Jiang and Doksum¹¹ can be used for estimating $\lambda(x)$.

Algorithm for estimating $\lambda(x)$:

Step 1: Pilot estimators of $\lambda(x)$: 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 $T / (8n_u^{5/8})$ as recommended by Muller & Wang,⁸ 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 our estimators (8), (13) and (14).

Step 2: Minimizing of $AMSE(b_n, x)$: Choose an equispaced grid of $m1$ points $\tilde{x}_i, i=1, \dots, m1$ between 0 and T . For each of the grid points \tilde{x}_i compute $AMSE(b_n, \tilde{x}_i)$ in (17) with $k=0$ and obtain its minimizer $\hat{b}(\tilde{x}_i)$ on the interval $[b_0/4, 4b_0]$, say. The minimization of $AMSE(b_n, \tilde{x}_i)$ may be computed via Discretisation.

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

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

where $w_i = K(\frac{\tilde{x}_i - x_r}{\hat{b}_0}) [M_{n,2} - M_{n,1}(\tilde{x}_i - x_r)]$ and

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

Step 4: Final hazard function estimators: Using (8), (13) and (14), obtain the estimators $\hat{\lambda}(x_r)$ by employing the bandwidth $\hat{b}(x_r)$, for $r=1, \dots, m2$.

The above algorithm can 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 can also be obtained via maximum likelihood if one has a plausible parametric model in mind.¹⁸

Appendix proofs of theorems

Proofs of theorems 3.1 and 3.2

(i) Using (1) and (9), we obtain that

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

$$= \int \frac{1}{b_n} K\left(\frac{u-x_0}{b_n}\right) \left(\frac{u-x_0}{b_n}\right)^\ell d\Lambda(u) + \int \frac{1}{b_n} K\left(\frac{u-x_0}{b_n}\right) \left(\frac{u-x_0}{b_n}\right)^\ell d[\Lambda_n(u) - \Lambda(u)] \\ \equiv \beta_{n\ell}(x_0) + \gamma_{n\ell}. \tag{18}$$

Let $\beta_n(x_0) = (\beta_{n0}, \dots, \beta_{np})^T$ and $\gamma_n(x_0) = (\gamma_{n0}, \dots, \gamma_{np})^T$. Then $S_n(x_0) = \beta_n(x_0) + \gamma_n(x_0)$. By (8), we have

$$\mathbf{B}[\hat{\mathbf{a}}(x_0) - \mathbf{a}(x_0)] = [\mathbf{S}^{-1}\beta_n(x_0) - \mathbf{B}\mathbf{a}(x_0)] + \mathbf{S}^{-1}\gamma_n(x_0) \\ \equiv \alpha_n(x_0) + \mathbf{S}^{-1}\gamma_n(x_0). \tag{19}$$

We will show that $\alpha_n(x_0)$ contributes to the bias term, and $\mathbf{S}^{-1}\gamma_n(x_0)$ the variance term of our estimator.

(a) By (18), a change of variable for integration, and the Taylor expansion, we get

$$\beta_{n\ell}(x_0) = \int K(t) t^\ell \lambda(x_0 + b_n t) dt \\ = \sum_{j=0}^p \frac{b_n^j}{j!} \lambda^{(j)}(x_0) s_{\ell+j} + \frac{b_n^{p+1}}{(p+1)!} \lambda^{(p+1)}(x_0) s_{\ell+p+1} (1 + o_p(1)).$$

Then

$$\alpha_n(x_0) = \frac{b_n^{p+1}}{(p+1)!} \lambda^{(p+1)}(x_0) \mathbf{S}^{-1} \mathbf{c}_p (1 + o(1)). \tag{20}$$

In particular, if $\lambda(x)$ is a polynomial up to order p in a neighborhood of x_0 , then the exact bias $\alpha_n(x_0)$ of the LP estimator is zero.

For $a_G < a_W$, by Gurler & Wang,¹ we have

$$\Lambda_n(x) - \Lambda(x) = \frac{1}{n} \sum_{i=1}^n \xi(Y_i, T_i, \delta_i, x) + r_n(x), \tag{21}$$

where for $x \geq 0$, any $b < b_W$ and $\delta = 1$ or 0 ,

$$\sup_{0 \leq x \leq b} |r_n(x)| = O\left(\frac{\log n}{n}\right), \text{ a.s. and}$$

$$\xi(y, t, \delta, x) = I(y \leq x, \delta = 1) / C(y) - \int_0^x I(t \leq u \leq y) [C(u)]^{-2} dW_1(u),$$

which satisfies that $E\xi(X_i, \delta_i, x) = 0$, and

$$\text{Cov}(\xi(Y_i, T_i, \delta_i, s), \xi(Y_i, T_i, \delta_i, t)) = \mathbf{g}(\min(s, t)),$$

where $\mathbf{g}(x) = \int_0^x [C(u)]^{-2} dW_1(u)$. Let $K_\ell(t) = K(t)t^\ell$. Using the definition of $\gamma_{n\ell}$, (21) and integration by parts, we obtain the following almost surely representation of $\gamma_{n\ell}(x_0)$:

$$\gamma_{n\ell}(x_0) = \sigma_{n\ell}(x_0) + e_{n\ell}(x_0), \tag{22}$$

where

$$\sigma_{n\ell}(x_0) = n^{-1} \sum_{i=1}^n \int \frac{1}{b_n} K\left(\frac{u-x_0}{b_n}\right) \left(\frac{u-x_0}{b_n}\right)^\ell d\xi(Y_i, T_i, \delta_i, u)$$

$$= (nb_n)^{-1} \sum_{i=1}^n \int \xi(Y_i, T_i, \delta_i, x_0 + b_n t) dK_\ell(t)$$

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

$$\sup_{0 \leq x_0 \leq b} |e_{n\ell}(x_0)| = O\left(\frac{\log n}{n}\right), \text{ a.s.} \tag{23}$$

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

$$\begin{aligned} \text{Cov}(\sigma_{n\ell}(x_0), \sigma_{nm}(x_0)) &= \frac{1}{nb_n} \int \int \frac{1}{b_n} g(\min(x_0 + b_n u, x_0 + b_n v)) dK_\ell(u) dK_m(v) \\ &= \frac{1}{nb_n^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_0 + b_n v K_\ell\left(\frac{u - x_0}{b_n}\right) dg(u) dK_m(v) \\ &= \frac{1}{nb_n} \frac{\lambda(x_0)}{C(x_0)} v_{\ell+m} (1 + o(1)). \end{aligned} \tag{24}$$

Let $\sigma_n(x_0) = (\sigma_{n0}, \dots, \sigma_{np})^T$ and $\mathbf{e}_n(x_0) = (e_{n0}, \dots, e_{np})^T$. Then by (22) and (24)

$$\sqrt{nb_n} \mathbf{S}^{-1} \gamma_n(x_0) = \sqrt{nb_n} \mathbf{S}^{-1} \sigma_n(x_0) + \sqrt{nb_n} \mathbf{S}^{-1} \mathbf{e}_n(x_0) \tag{25}$$

and $\text{Cov}(\sigma_n(x_0), \sigma_n(x_0)) = \frac{1}{nb_n} \frac{\lambda(x_0)}{C(x_0)} \mathbf{V}^* (1 + o(1))$. By the central limit theorem, we get

$$\sqrt{nb_n} \sigma_n(x_0) \xrightarrow{L} N\left(0, \frac{\lambda(x_0)}{C(x_0)} \mathbf{V}^*\right). \tag{26}$$

By (23), we know

$$\sup_{0 \leq x_0 \leq b} |\mathbf{S}^{-1} \mathbf{e}_n(x_0)| = O\left(\frac{\log n}{n}\right), \text{ a.s.} \tag{27}$$

Then by (25), (26) and (27)

$$\sqrt{nb_n} \mathbf{S}^{-1} \gamma_n(x_0) \xrightarrow{L} N\left(0, \frac{\lambda(x_0)}{C(x_0)} \mathbf{S}^{-1} \mathbf{V}^* \mathbf{S}^{-1}\right). \tag{28}$$

Combination of (19), (20) and (28) completes the proof of Theorem 3.2.

(ii) Note that $\sigma_{n\ell}(x_0)$, for $\ell = 0, \dots, p$, are i.i.d.'s sums. By the SLN, we know $\sigma_{n\ell}(x_0) \xrightarrow{a.s.} E(\sigma_{n\ell}(x_0)) = 0$. Then $\sigma_n(x_0) \xrightarrow{a.s.} \mathbf{0}$. This combined with (22) and (23) yields $\gamma_n(x_0) \xrightarrow{a.s.} \mathbf{0}$. Therefore, by (19) and (20),

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

Proofs of Theorem 3.3 The result follows by the same argument as in part (i) of the proof of Theorem 3.2.

Acknowledgements

This work is partially supported by Natural Science Foundation of Fujian Providence of China (2016J01024) and by Education and Scientific Research Projects for young and middle-aged teachers of Fujian Providence (JAT160383).

Conflicts of interest

Author declares that there are no conflicts of interest.

References

- Gürler Ü, Wang JL. Nonparametric estimation of hazard functions and their derivatives under truncation model. *Ann Inst Statist Math.* 1993;(45):249–264.
- Gross ST, Lai TL. Nonparametric estimation and regression analysis with left-truncated and right censored data. *Jour Amer Statist Assoc.* 1996;(91):1166–1180.
- Uzunogullari U, Wang JL. A comparison of hazard rate estimators for left truncated and right censored data. *Biometrika.* 1992;79(2):297–310.
- Wu SS, Wells MT. Nonparametric estimation of hazard functions by wavelet methods. *Journal of Nonparametric Statistics.* 2000;15(2):187–203.
- Fan J, Gijbels I. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *JR Statist Soc B.* 1995;(57):371–394.
- Woodrooffe M. Estimating a distribution function with truncated data. *Ann Statist.* 1985;(13):163–177.
- Muller HG, Wang JL. Locally adaptive hazard smoothing. *Prob Th Rel Fields.* 1989;85(4):523–538.
- Muller HG, Wang JL. Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics.* 1994;50(1):61–76.
- Marron JS, Wand MP. Exact mean integrated squared error. *Ann Statist.* 1992;20(2):712–736.
- Hall P, Wehrly TE. A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Jour Amer Statist Soc.* 1991;86(415):665–672.
- Jiang J, K Doksum. On Local Polynomial Estimation of Hazard Functions and Their Derivatives under Random Censoring. *Lecture Notes–Monograph Series.* 2003;(42):463–481.
- Jones MC. Simple boundary correction for kernel density estimation. *Statistics and Computing.* 1993;3(3):135–146.
- Lejeune M, Sarda P. Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis.* 1992;(14):457–471.
- Hess KR, Serachitopol DM, Brown BW. Hazard function estimators: a simulation study. *Statistics in Medicine.* 1999;18(22):3075–3088.
- Fan J, Gijbels I. *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall; 1996.
- Patil PN. On the least squares cross-validation bandwidth in hazard rate estimation. *Ann Statist.* 1993;21(4):1792–1810.
- Gonzalez-Manteiga W, Cao R, Marron JS. Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Jour Amer Statist Assoc.* 1996;91(435):1130–1140.
- Tsai WY, Jewell NP, Wang MC. A note on the product-limit estimator under right censoring and left truncation. *Biometrika.* 1987;74(4):883–886.