Journal of the Chinese
Statistical Association
Vol. 47, (2009) 19–38

LOCAL POLYNOMIAL ESTIMATION OF HAZARD RATES UNDER RANDOM CENSORING

Ming-Yen Cheng[1,2] and Chun-Yi Lee[2]

[1]**Department of Statistical Science, University College London**

[2]**Department of Mathematics, National Taiwan University**

ABSTRACT

In many survival studies, observation on the occurrence of the event of interest (called a failure) may be prevented by the previous occurrence of another event (called a censoring event). We assume the random censorship model in which the censoring time is independent of the survival time. Considering least squares local linear and local quadratic approximations to the Nelson-Aalen estimator of the cumulative hazard function, the estimators of the linear coefficients are called the local linear and local quadratic estimators of the hazard rate, respectively. The asymptotic normal distributions of the local linear and local quadratic estimators are established. Performance of the proposed estimators is illustrated by simulations. We compare the proposed estimators with the kernel estimator and the Jiang and Doksum (2003) estimator by means of the estimated MISE, and find that the local quadratic estimator behaves favorably.

Key words and phrases: Censoring, cumulative hazard, hazard rate, kernel method, local polynomial estimation.
JEL classification: C13, C14, C24.

1. Introduction

In many survival studies, observation on the occurrence of the event of interest (called a failure) may be prevented by the previous occurrence of another event (called a censoring event). Let $T_1, \ldots, T_n$ be the survival times for the n subjects under study, and $C_1, \ldots, C_n$ be the corresponding censoring times. We can only observe $X_i = \min(T_i, C_i)$ and $\delta_i = I(X_i = T_i)$ (the censoring indicator), $i = 1, \ldots, n$, where $I(\cdot)$ is the indicator function. Here, we assume the random censorship model, i.e. C_i and T_i are independent, and consider the problem of nonparametrically estimating the underlying hazard function.

Tanner and Wong (1983) studied asymptotic properties of kernel estimators of the hazard function based on the idea of convolution using Hajék projection. The kernel estimator can be regarded as a convolution of the derivative of the Nelson-Aalen estimator, which is the empirical cumulative hazard, with a kernel function. Müller and Wang (1990) considered local bandwidth choice for 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. Jiang and Doksum (2003) considered local polynomial estimators of hazard functions and their derivatives: the Dirac function is used to define a generalized empirical hazard rate, denoted as $\widetilde{\lambda}_n(\cdot)$, whose integration up to time x equals to the Nelson-Aalen estimator evaluated at x. The resulting estimators automatically correct boundary effects.

The empirical hazard rate defined in Jiang and Doksum (2003) exists only in the space of Schwartz distributions and is not directly computable. Here, instead of using the Dirac function to define a generalized empirical hazard rate, we consider a local polynomial approximation to the Nelson-Aalen estimator of the cumulative hazard function directly and then estimate its derivative (i.e. the hazard) by the derivative of the fitted local polynomial. Cheng et al. (2006) suggested local polynomial approximation to an estimator of the cumulative hazard function which is taken as a transformation of the Kaplan-Meier estimator of the cumulative distribution function. Both of the above methods automatically correct the boundary effects. However, our method is simpler since it does not involve transformation of the Kaplan-Meier estimator. Loader (1999) introduced local likelihood estimators of hazard rates, which employ local exponential polynomial approximation to the Kaplan-Meier estimator. Unlike our method, this approach does not admit explicit solution to the maximization of the local likelihood and requires numerical solutions.

In Section 2, we review the kernel estimator and the estimator of Jiang and Doksum (2003) before introducing our local polynomial estimators. The asymptotic normal distributions of the proposed estimators are given in Section 3. Numerical illustration is given in Section 4. Section 5 provides some discussion. The proofs for the asymptotic properties are deferred to Section 6.

2. Hazard Estimation

Under the assumption of the random censorship, let $T_1, \ldots, T_n$ be i.i.d. with distribution function F, independent of $C_1, \ldots, C_n$ which are i.i.d. with distribution function G. Let L and l denote the distribution function and the density of X_i, respectively. Then, $\bar{L} = \bar{F}\bar{G}$, where for any distribution function E, $\bar{E} = 1 - E$ is the corresponding survival function. Further let $\Lambda(x) = -\log(\bar{F}(x))$ be the cumulative hazard function. What we are interested in is to nonparametrically estimate the hazard function $\lambda(x) = \Lambda'(x) = f(x)/\bar{F}(x)$ on an interval $[0, T]$ such that $L(T) < 1$. Here we assume that the density $f = F'$ exists.

In order to introduce the estimators, we define the following notations.

(1) $L_1(x) = \mathrm{P}(X_i \leq x, \delta_i = 1)$ is the distribution function for the uncensored observations.

(2) $L_{1n}(x) = \sum_{i=1}^n I(X_i \leq x, \delta_i = 1)/n$ is the empirical distribution function based on the uncensored observations which can be used to estimate $L_1(x)$.

(3) $L_n(x) = \sum_{i=1}^n I(X_i \leq x)/n$ is the empirical distribution function based on the all the X's and can be used to estimate $L(x)$.

Note that, under the random censoring model, we have

$$\lambda(x) = \frac{f(x)}{\bar{F}(x)} = \frac{\bar{G}(x)\, dF(x)/dx}{\bar{F}(x)\bar{G}(x)} = \frac{dL_1(x)/dx}{\bar{L}(x)}\,.$$

The Nelson-Aalen estimator of $\Lambda(x)$ is

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

where $X_{(i)}$ is the order statistic corresponding to X_i, and $\delta_{(i)}$ is the corresponding indicator variable. Tanner and Wong (1983) proposed the following kernel estimator of the hazard $\lambda(x)$:

$$\widehat{\lambda}(x) = \int K_h(u-x)\, d\Lambda_n(u) = \sum_{i=1}^{n} K_h\left(X_{(i)} - x\right) \frac{\delta_{(i)}}{n-i+1},$$

where K is a kernel function, h is a bandwidth and $K_h(\cdot) = K(\cdot/h)/h$. The kernel estimator is a convolution of the derivative of the Nelson-Aalen estimator Λ_n with an appropriate kernel function K_h. Here, the derivative of Λ_n can be expressed by

$$\lambda_n(x) = \sum_{i=1}^{n} \frac{\delta_{(i)}}{n-i+1} I\left(x = X_{(i)}\right).$$

where $I\left(x = X_{(i)}\right)$ is the indicator variable.

Jiang and Doksum (2003) defined the following generalized empirical hazard rate:

$$\widetilde{\lambda}_n(x) = \sum_{i=1}^{n} \frac{\delta_{(i)}}{n-i+1} D\left(x - X_{(i)}\right),$$

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 $\int_0^x \widetilde{\lambda}_n(t)\, dt = \Lambda_n(x)$, however $\widetilde{\lambda}_n(\cdot)$ exists only in the space of Schwartz distributions and is not computable. They considered the following local least squares problem: for $p = 0, 1, 2, \ldots,$

$$\min_{a_0, a_1, \cdots, a_p} \int_0^\infty K_h\left(X_{(i)} - x\right) \left(\lambda_n(u) - \sum_{j=0}^{p} a_j (u-x)^j\right)^2 du.$$

Then their estimator of $\lambda(x)$ is defined as the the fitted value of a_0. When x is an interior point and $p = 1$, the Jiang and Doksum estimator of the hazard rate is the same as the kernel estimator.

Suppose the cumulative hazard function has a p-th order derivative at x. Then, using a Taylor expansion, locally it can be approximated by a p-th order polynomial:

$$\Lambda(u) \approx \sum_{j=0}^{p} \frac{\Lambda^{(j)}(x)}{j!}(u-x)^j = \sum_{j=0}^{p} a_j(x)(u-x)^j$$

for u in a neighborhood of x. Note that $a_j(x) = \Lambda^{(j)}(x)/j!$, $j = 0, 1, 2, \ldots, p$. Let us consider the following local least squares problem: for $p = 0, 1, 2$,

$$\min_{a_0, a_1, \cdots, a_p} \sum_{i=1}^{n} K_h\left(X_{(i)} - x\right)\left(\Lambda_n\left(X_{(i)}\right) - \sum_{j=0}^{p} a_j\left(X_{(i)} - x\right)^j\right)^2 .$$

Denote the solution of $(a_0, a_1, \ldots, a_p)^T$ to the above local least squares problem by $\widehat{a}(x) \equiv (\widehat{a}_0(x), \widehat{a}_1(x), \ldots, \widehat{a}_p(x))^T$. Then, our estimator of the hazard rate $\lambda(x) = \Lambda'(x)$ is $\widehat{a}_1(x)$. We consider two cases: when $p = 1$ the above procedure yields the local linear estimator $\widehat{\lambda}_{loclin}(x)$, and when $p = 2$, we have the local quadratic estimator $\widehat{\lambda}_{locqua}(x)$. Denote

$$S_k(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(X_{(i)} - x\right)\left(X_{(i)} - x\right)^k, \quad k = 0, 1, 2, 3, 4,$$

$$T_k(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(X_{(i)} - x\right)\left(X_{(i)} - x\right)^k \sum_{X_{(j)} \le X_{(i)}} \frac{\delta_{(j)}}{n-j+1}, \quad k = 0, 1, 2.$$

Then, we can write

$$\widehat{\lambda}_{loclin}(x) = \frac{S_0(x)T_1(x) - S_1(x)T_0(x)}{S_0(x)S_2(x) - S_1^2(x)}, \quad \text{and} \quad \widehat{\lambda}_{locqua}(x) = \frac{\Delta_1(x)}{\Delta(x)},$$

where

$$\begin{aligned}\Delta(x) = {} & S_0(x)S_2(x)S_4(x) + 2S_1(x)S_2(x)S_3(x) \\ & - S_2^3(x) - S_0(x)S_3^2(x) - S_1^2(x)S_4(x),\end{aligned}$$

and

$$\begin{aligned}\Delta_1(x) = {} & [S_2(x)S_3(x) - S_1(x)S_4(x)]\, T_0(x) + \left[S_0(x)S_4(x) - S_2^2(x)\right] T_1(x) \\ & + [(S_1(x)S_2(x) - S_0(x)S_3(x)]\, T_2(x).\end{aligned}$$

3. Asymptotic properties

For a given point $x_0 \in (0, T)$, the following assumptions are needed in order to investigate the asymptotic normal distributions of $\widehat{\lambda}_{loclin}(x_0)$ and $\widehat{\lambda}_{locqua}(x_0)$.

(A1) The hazard function $\lambda(x)$ has a continuous third derivative at the point x_0 .

(A2) The sequence of bandwidths h tends to zero such that $nh \to \infty$ as $n \to \infty$.

(A3) $L(x)$ has a continuous fifth derivative at the point x_0 .

(A4) The kernel function K is continuous, symmetric and of bounded variation, and it has a bounded support $[-1, 1]$. Assume that $\mu_k = \int K(u)u^k\, du$, $k = 0, 2, 4, 6$ exist , and $v_k = \int K^2(u)u^k\, du$, $k = 0, 1, 2, 3, 4$ exist.

In this section, we will establish the asymptotic normality of the local linear and local quadratic estimators. Since both of our estimators are combinations of $S_k(x_0)$ and $T_k(x_0)$, we will analyze $S_k(x_0)$ and $T_k(x_0)$ for a given point $x_0 \in (0, T)$ first.

Lemma 3.1. *Under conditions* (A2) – (A4),

$$S_k(x_0) = \begin{cases} h^k \mu_k l(x_0) + \frac{1}{2} h^{k+2} \mu_{k+2} l''(x_0) + o_p(h^{k+2})\,, & k = 0, 2, 4\,; \\ h^{k+1} \mu_{k+1} l'(x_0) + \frac{1}{6} h^{k+3} \mu_{k+3} l'''(x_0) + o_p(h^{k+3})\,, & k = 1, 3\,. \end{cases}$$

Proof. Notice that $S_k(x_0)$ are sums of i.i.d. random variables, for $k = 0, 1, 2$. By the Strong Law of Large Numbers, we know that $S_k(x_0) \xrightarrow{a.s.} \mathrm{E}(S_k(x_0))$. Then, using change of variable for integration, a Taylor expansion and the assumption that the kernel function is symmetric, we have the following result:

$$\begin{aligned} S_k(x_0) &= \frac{1}{n} \sum_{i=1}^{n} K_h(X_{(i)} - x_0)(X_{(i)} - x_0)^k \\ &= \int K_h(u - x_0)(u - x_0)^k l(u)\, du\, (1 + o_p(1)) \\ &= h^k \int K(v) v^k l(x_0 + hv)\, dv\, (1 + o_p(1)) \\ &= \begin{cases} h^k \mu_k l(x_0) + \frac{1}{2} h^{k+2} \mu_{k+2} l''(x_0) + o_p(h^{k+2})\,, & k = 0, 2, 4\,, \\ h^{k+1} \mu_{k+1} l'(x_0) + \frac{1}{6} h^{k+3} \mu_{k+3} l'''(x_0) + o_p(h^{k+3})\,, & k = 1, 3\,. \end{cases} \end{aligned}$$

□

Theorem 3.1. *Under conditions* (A1) – (A4)*, for* $k = 0, 1, 2$*,*

$$\sqrt{nh^{-2k-1}}\left\{T_k(x_0) - \widetilde{\beta_0^{k1}}(x_0)\right\} \xrightarrow{d} \mathcal{N}\Big(0, L(x_0)l(x_0)g(x_0)v_{2k}\Big),$$

where

$$\widetilde{\beta_0^{k1}}(x_0)$$

$$= \begin{cases} h^k\mu_k\Lambda(x_0)l(x_0) \\ \quad +h^{k+2}\mu_{k+2}\left[\lambda(x_0)l'(x_0) + \frac{1}{2}\Lambda(x_0)l''(x_0) + \frac{1}{2}\lambda'(x_0)l(x_0)\right] \\ \qquad +h^{k+4}\mu_{k+4}\left\{\frac{1}{4}\lambda'(x_0)l''(x_0) + \frac{1}{6}\left[\lambda(x_0)l'''(x_0) + \lambda''(x_0)l(x_0)\right]\right. \\ \qquad\qquad \left. +\frac{1}{24}\left[\Lambda(x_0)l''''(x_0) + \lambda'''(x_0)l(x_0)\right]\right\}, \\ & k = 0, 2, \\ h^{k+1}\mu_{k+1}\left[\Lambda(x_0)l'(x_0) + \lambda(x_0)l(x_0)\right] \\ \quad +h^{k+3}\mu_{k+3}\left\{\frac{1}{2}\left[\lambda(x_0)l''(x_0) + \lambda'(x_0)l'(x_0)\right]\right. \\ \qquad\qquad \left. +\frac{1}{6}\left[\Lambda(x_0)l'''(x_0) + \lambda''(x_0)l(x_0)\right]\right\}, \\ & k = 1. \end{cases}$$

and

$$g(x_0) = \int_0^{x_0} [\bar{L}(u)]^{-2}\, dL_1(u).$$

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

$$\sqrt{nh}\left\{\widehat{\lambda}_{loclin}(x_0) - \lambda(x_0) - \frac{h^2}{6}\frac{3(\mu_4 - \mu_2^2)\lambda'(x_0)l'(x_0)/l(x_0) + \mu_4\lambda''(x_0)}{\mu_2}\right\}$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{L(x_0)g(x_0)}{\mu_2^2 l(x_0)}v_2\right).$$

$$\sqrt{nh}\left\{\widehat{\lambda}_{locqua}(x_0) - \lambda(x_0) - \frac{h^2}{6}\frac{\mu_4}{\mu_2}\lambda''(x_0)\right\} \xrightarrow{d} \mathcal{N}\left(0, \frac{L(x_0)g(x_0)}{\mu_2^2 l(x_0)}v_2\right).$$

The proofs of Theorem 3.1 and Theorem 3.2 are postponed to Section 6.

4. Simulation

In this section, we compare the performance of our estimators with the kernel estimator and the Jiang and Doksum's estimator (J&D for short) ($p = 1$) through simulation studies. We use the Epanechnikov kernel $K(x) = 0.75(1-x^2)I(|x| \leq 1)$ for all the estimators. In our simulation studies, we consider two underlying distributions, Uniform model and Weibull model.

In order to control the amount of censoring, we use the proportional censorship model (see Koziol and Green, 1976; González-Manteiga et al., 1996), in which G is defined by $\bar{G} = \bar{F}^\eta$ for some $\eta > 0$. This model gives a probability of censoring $\xi = \eta/(1+\eta)$, thus allows a simple control of the amount of censoring in the simulation through the choice of η.

Example 4.1 (Uniform) We simulate 500 samples with size $n = 500$ from the uniform distribution: $T_j \overset{iid}{\sim} F(x) = \text{Uniform}[0,1]$. Then the true hazard rate is $\lambda(x) = 1/(1-x)$ which is a strictly increasing function on $[0,1]$ and has the range $[0,+\infty)$, see the solid line in Figure 1. The proportional censorship model is used and then, $C_j \overset{iid}{\sim} G(x) = 1-\bar{F}^\eta(x)$. We choose $\eta = \frac{1}{9}$ and $\frac{1}{2}$ so that the probabilities for censoring are $\xi = \frac{1}{10}$ and $\frac{1}{3}$, respectively. Besides, we make the assumption that T_j is independent of C_j. Figure 1 shows our local linear estimates, local quadratic estimate, the kernel estimate, and the J&D estimate with $p = 1$, which is the same as the kernel estimator for interior points, based on one of the 500 simulated samples. The four estimators behave similarly at the interior points but very differently especially at the right boundary points $x \in (1-h, 1]$. The kernel estimator suffers from boundary effects near the endpoints. We use a fixed bandwidth $h = 0.05$ for all the four estimators.

Example 4.2 (Weibull) We simulate 500 samples with size $n = 500$ from the Weibull distribution: $T_j \overset{iid}{\sim} F(x) = 1 - \exp(-\sqrt{x})$, and let $C_j \overset{iid}{\sim} G(x) = 1 - \bar{F}^\eta(x)$ for $\eta = \frac{1}{9}$ and $\frac{1}{2}$. T_j is independent of C_j. Then the hazard rate is a strictly decreasing function on $[0,1]$ and has the range $[0,+\infty)$, see the solid line in Figure 2. Figure 2 depicts the four estimates based on one of the 500 simulated samples. It shows that all of the four estimators behave differently at the boundary points $x \in [0,h)$. Among these four estimators, the kernel estimator performs the worst. Here, we use a fixed bandwidth $h = 0.1$ for the four estimators.

To examine the four estimators more closely, we compute the estimated MISEs.

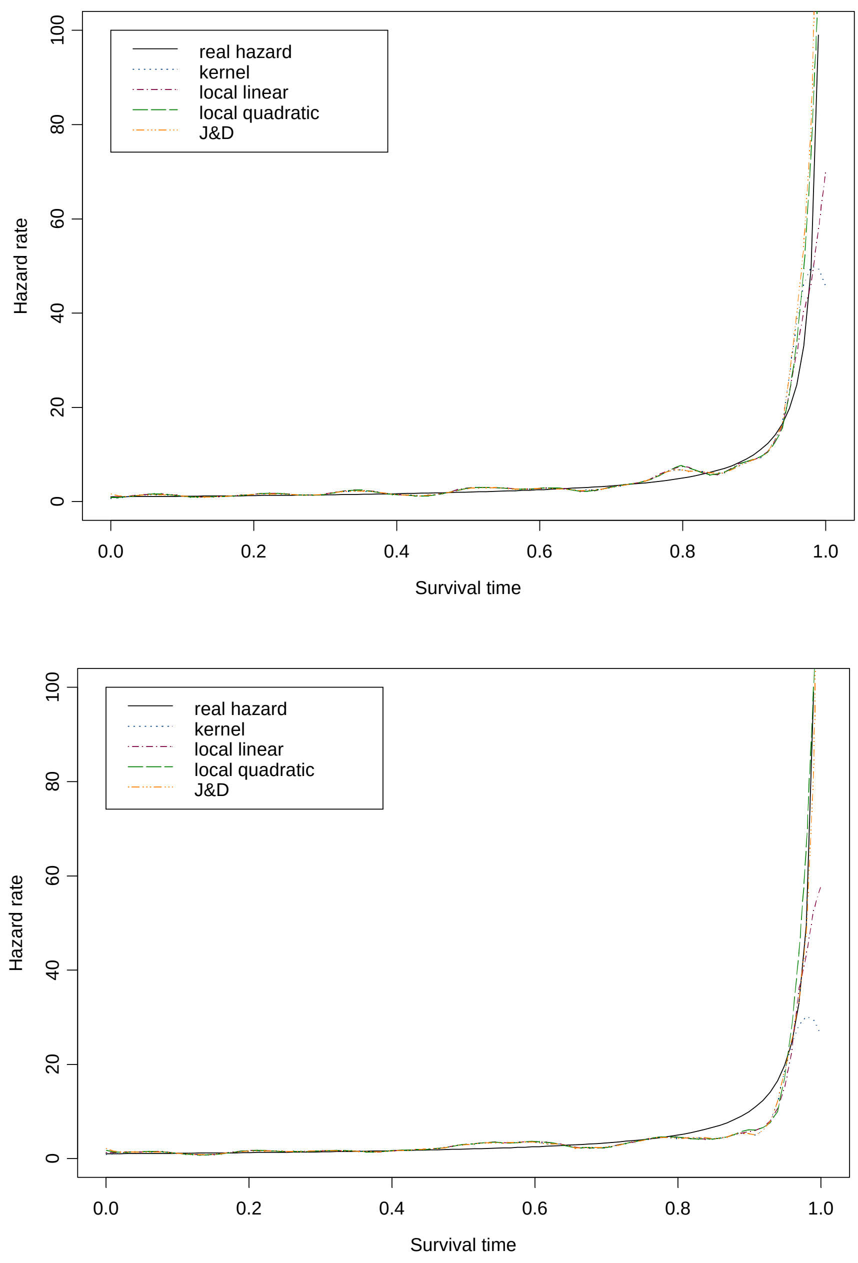

Figure 1 Simulation results for Example 4.1. The probability for censoring ξ is $\frac{1}{10}$ in the top plot and is $\frac{1}{3}$ in the bottom plot.

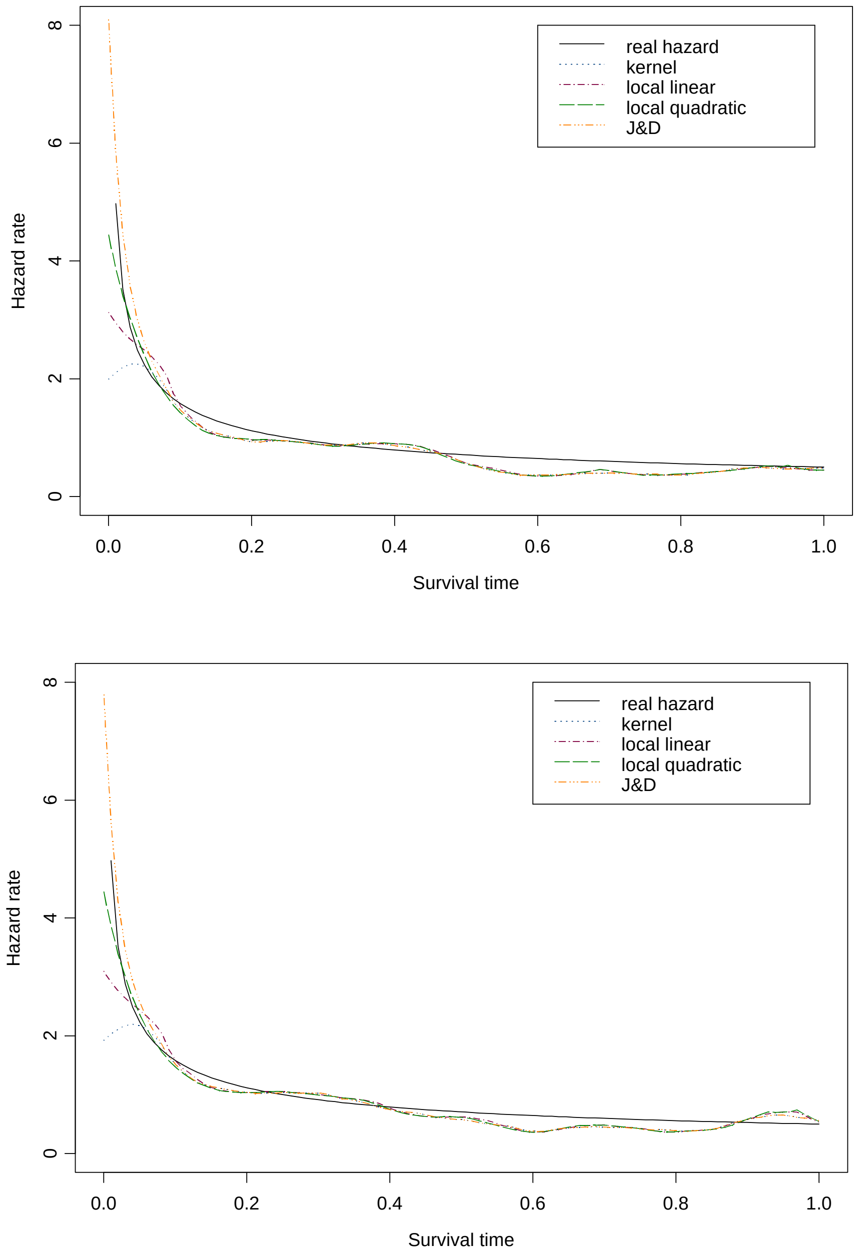

Figure 2 Simulation results for Example 4.2. The probability for censoring ξ is $\frac{1}{10}$ in the top plot and is $\frac{1}{3}$ in the bottom plot.

We choose h to be a sequence of $0.08 * 1.1^k$, $k = 0, 1, \ldots, 15$ and find the smallest estimated MSE at every grid point among these h. Then the estimated MISE is a numeric integration of the smallest estimated MSE. It provides a measure of overall performance when an ideal local bandwidth is used. Because there are few observations near the endpoint 1 in the uniform model, we consider to estimate MISE form 0 to the maximum uncensored data in each simulation. Let $\widetilde{X}_{(i)}^{(j)}$ be the i^{th} order statistic of the uncensored observations in the j^{th} simulation. We consider $a = \min_j \left(\max_i \left(\widetilde{X}_{(i)}^{(j)} \right) \right)$ and estimate $\text{MISE} = \text{E}\left\{ \int_0^a [\hat{\lambda}(x) - \lambda(x)]^2 \, dx \right\}$ in which $a = 0.98$ and 0.92 for the $\frac{1}{10}$ and $\frac{1}{3}$ censoring rates, respectively.

Table 1 presents the estimated MISE's for the four estimators under the uniform model. It shows that the local quadratic estimator has the smallest estimated MISE among these estimators for the 500 samples under $\frac{1}{10}$ censoring rate. Notice that the true hazard in the uniform case is $\lambda(x) = 1/(1-x)$ which tends to $+\infty$ when x gets close to 1. The J&D estimator tends to over-estimate the real hazard and hence its MSE's at right boundary points ($x \in (1-h, 0.98)$) are much larger than that of our local quadratic estimator under $\frac{1}{10}$ censoring rate. The local linear estimator has the smallest estimated MISE among these estimators for the 500 samples under $\frac{1}{3}$ censoring rate.

Table 2 reports the estimated MISE's for these estimators under the weibull model. It shows that the local quadratic estimator behaves favorably compared to the other estimators under both censoring rates. We conclude from both the numerical and the graphical viewpoints that the performance of our local quadratic estimator is superior to the J&D estimator.

5. Discussion

We propose local polynomial estimation of hazard rate based on a local polynomial approximation to the Nelson-Aalen estimator of the cumulative hazard function using the local least squares idea. The proposed estimators admit explicit forms, hence can

Table 1 Estimated MISE under uniform model

	local quadratic	local linear	J&D	kernel
$\xi = 1/10$	1.10722	3.49005	9.37975	3.62497
$\xi = 1/3$	0.45119	0.18086	0.93186	0.37788

Table 2 Estimated MISE under weibull model

	local quadratic	local linear	J&D	kernel
$\xi = 1/10$	0.02026	0.04317	0.02912	0.08261
$\xi = 1/3$	0.02360	0.04651	0.03091	0.08589

be implemented easily. In addition, they are free from boundary effects which is a serious problem of the kernel estimator suffer in practice. Compared to the the local polynomial methods proposed by Jiang and Doksum (2003) and Cheng et al. (2006), our estimators are easier to compute while processing the same theoretical properties. By contrast, the local likelihood approach of Loader (1999) requires numerical solutions to the maximum local likelihood problem and asymptotic normal distributions of the local likelihood estimators remain unknown. On the other hand, it guarantees that the hazard rate estimator is always nonnegative, while the local polynomial approaches do not. It is an interesting topic for future studies to investigate theoretical properties of the local likelihood hazard estimators and to compare it with the local polynomial methods through numerical studies.

6. Proofs

Proof of Theorem 3.1

1. By the definitions of $T_k(x_0)$ and the Nelson-Aalen estimator, and using change of

variable, $T_k(x_0)$ can be decomposed into four components:

$$
\begin{aligned}
T_k(x_0) &= \frac{1}{n}\sum K_h(X_{(i)}-x_0)(X_{(i)}-x_0)^k \sum_{X_{(j)}\le X_{(i)}} \frac{\delta_{(j)}}{n-j+1} \\
&= \int K_h(u-x_0)(u-x_0)^k \sum_{X_{(j)}\le u} \frac{\delta_{(j)}}{n-j+1}\, dL_n(u) \\
&= h^k \int K(v)v^k \left(\int_0^{x_0+hv} d\Lambda_n(t)\right) dL_n(x_0+hv) \\
&= h^k \int K(v)v^k \beta_{nv}(x_0)\, dL(x_0+hv) \\
&\quad +\, h^k \int K(v)v^k \gamma_{nv}(x_0)\, dL(x_0+hv) \\
&\quad +\, h^k \int K(v)v^k \beta_{nv}(x_0)\,(dL_n(x_0+vh)-dL(x_0+hv)) \\
&\quad +\, h^k \int K(v)v^k \gamma_{nv}(x_0)\,(dL_n(x_0+vh)-dL(x_0+hv)) \\
&\equiv \widetilde{\beta_n^{k1}}(x_0)+\widetilde{\gamma_n^{k1}}(x_0)+\widetilde{\beta_n^{k2}}(x_0)+\widetilde{\gamma_n^{k2}}(x_0)\,, \quad k=0,1,2\,.
\end{aligned}
$$

where

$$
\begin{aligned}
\int_0^{x_0+hv} d\Lambda_n(t) &= \int_0^{x_0+hv} d\Lambda(t) + \int_0^{x_0+hv} (d\Lambda_n(t)-d\Lambda(t)) \\
&\equiv \beta_{nv}(x_0)+\gamma_{nv}(x_0)\,.
\end{aligned}
$$

We will show that $\widetilde{\beta_n^{k1}}(x_0)$ represents the expectation of $T_k(x_0)$, $\widetilde{\gamma_n^{k1}}(x_0)$ is the random error, and $\widetilde{\beta_n^{k2}}(x_0)$ and $\widetilde{\gamma_n^{k2}}(x_0)$ are the remainder terms.

2. Before looking into these four components, we analyze $\beta_{nv}(x_0)$ and $\gamma_{nv}(x_0)$.

   (a) First, by a Taylor expansion, we get

$$
\begin{aligned}
\beta_{nv}(x_0) &= \int_0^{x_0+hv} d\Lambda(t) = \int_0^{x_0+hv} \lambda(t)\,dt = \Lambda(x_0+hv) \\
&= \Lambda(x_0)+hv\lambda(x_0)+\frac{1}{2}h^2v^2\lambda'(x_0)+o(h^2)\,.
\end{aligned}
$$

   (b) Second, from Lo et al. (1989), there is an asymptotic representation of the Nelson-Aalen estimator as a sum of i.i.d. random variables:

$$
\Lambda_n(x_0)-\Lambda(x_0) = \frac{1}{n}\sum_{i=1}^{n}\xi(X_i,\delta_i,x_0)+r_n(x_0)\,, \tag{1}
$$

where

$$\sup_{0\le x_0\le T} |r_n(x_0)| = O\left(\frac{\log n}{n}\right) \qquad \text{a.s.}$$

for $z \ge 0$, $x_0 \ge 0$, $\delta = 1$ or 0,

$$\xi(z,\delta,x_0) = -g(\min(z,x_0)) + \frac{I(z \le x_0, \delta = 1)}{\bar{L}(x_0)}.$$

Note that the random variable $\xi(X_i,\delta_i,x_0)$ are bounded, uniformly in $0 \le x_0 \le T$, $\mathrm{E}\,\xi(X_i,\delta_i,x) = 0$, and

$$\mathrm{Cov}(\xi(X_i,\delta_i,s),\xi(X_i,\delta_i,t)) = g(\min(s,t)).$$

Using the definition of $\gamma_{nv}(x_0)$ and the asymptotic representation of the Nelson-Aalen estimator (1), we can obtain the following almost surely representation of $\gamma_{nv}(x_0)$:

$$\gamma_{nv}(x_0) = \sigma_{nv}(x_0) + e_{nv}(x_0) \tag{2}$$

where

$$\sigma_{nv}(x_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{x_0+hv} d\xi(X_i,\delta_i,t) = \frac{1}{n}\sum_{i=1}^{n}\xi(X_i,\delta_i,x_0+hv)$$

is the stochastic component of $T_k(x_0)$, and $e_{nv}(x_0)$ is the negligible error of the approximation which satisfies

$$\sup_{0\le x_0\le T} |e_{nv}(x_0)| = O\left(\frac{\log n}{n}\right) \text{ a.s.}$$

3. Then, we study the four elements of $T_k(x_0)$.

   (a) By the definition of $\widetilde{\beta_n^{k1}}(x_0)$ and Taylor expansions, we have the following result:

$$\widetilde{\beta_n^{k1}}(x_0)$$
$$= \begin{cases} h^k\mu_k\Lambda(x_0)l(x_0) \\ \quad +h^{k+2}\mu_{k+2}\left[\lambda(x_0)l'(x_0) + \frac{1}{2}\Lambda(x_0)l''(x_0) + \frac{1}{2}\lambda'(x_0)l(x_0)\right] \\ \qquad +h^{k+4}\mu_{k+4}\left\{\frac{1}{4}\lambda'(x_0)l''(x_0) + \frac{1}{6}\left[\lambda(x_0)l'''(x_0) + \lambda''(x_0)l(x_0)\right]\right. \\ \qquad\qquad \left. +\frac{1}{24}\left[\Lambda(x_0)l''''(x_0) + \lambda'''(x_0)l(x_0)\right]\right\} + o(h^{k+4}), \\ \hfill k = 0,2; \\ h^{k+1}\mu_{k+1}\left[\Lambda(x_0)l'(x_0) + \lambda(x_0)l(x_0)\right] \\ \quad +h^{k+3}\mu_{k+3}\left\{\frac{1}{2}\left[\lambda(x_0)l''(x_0) + \lambda'(x_0)l'(x_0)\right]\right. \\ \qquad\qquad \left. +\frac{1}{6}\left[\Lambda(x_0)l'''(x_0) + \lambda''(x_0)l(x_0)\right]\right\} + o(h^{k+3}), \\ \hfill k = 1. \end{cases}$$

(b) Next, using the almost surely representation of $\gamma_{nv}(x_0)$ given in (2), $\widetilde{\gamma_n^{k1}}(x_0)$ can be decomposed into two parts as in the following lemma.

Lemma 6.1. *Under conditions* (A2) *and* (A4)*, for* $k = 0, 1, 2$,

$$\widetilde{\gamma_n^{k1}}(x_0) = \widetilde{\sigma_n^{k1}}(x_0) + \widetilde{e_n^{k1}}(x_0)\,,$$

where

$$\widetilde{\sigma_n^{k1}}(x_0) = \frac{1}{n}\sum_{i=1}^{n} h^k \int K(v)v^k \xi(X_i, \delta_i, x_0 + hv)\, dL(x_0 + hv)$$

is the stochastic component of $T_k(x_0)$*, and* $\widetilde{e_n^{k1}}(x_0)$ *is the remainder term of the approximation satisfying*

$$\sup_{0 \le x_0 \le T} \left|\widetilde{e_n^{k1}}(x_0)\right| = O\left(h^k\left(\frac{\log n}{n}\right)\right).$$

Proof.

$$\begin{aligned}
\widetilde{\gamma_n^{k1}}(x_0) &= h^k \int K(v)v^k \gamma_{nv}(x_0)\, dL(x_0 + hv) \\
&= h^k \int K(v)v^k \sigma_{nv}(x_0)\, dL(x_0 + hv) \\
&\quad + h^k \int K(v)v^k e_{nv}(x_0)\, dL(x_0 + hv) \\
&\equiv \widetilde{\sigma_n^{k1}}(x_0) + \widetilde{e_n^{k1}}(x_0)\,, \\
\widetilde{\sigma_n^{k1}}(x_0) &= h^k \int K(v)v^k \sigma_{nv}(x_0)\, dL(x_0 + hv) \\
&= \frac{1}{n}\sum_{i=1}^{n} h^k \int K(v)v^k \xi(X_i, \delta_i, x_0 + hv)\, dL(x_0 + hv)\,, \\
\widetilde{e_n^{k1}}(x_0) &= h^k \int K(v)v^k e_{nv}(x_0)\, dL(x_0 + hv) \\
&= O\left(h^k\left(\frac{\log n}{n}\right)\right).
\end{aligned}$$

$\square$

Lemma 6.2. *Under conditions* (A2) $-$ (A4)*, for* $k = 0, 1, 2$,

$$\sqrt{nh^{-2k-1}}\, \widetilde{\sigma_n^{k1}}(x_0) \xrightarrow{d} \mathcal{N}\Big(0, L(x_0)l(x_0)g(x_0)v_{2k}\Big).$$

Proof. Using the fact that $\mathrm{E}\,\xi(X_i,\delta_i,x_0)=0$, we can easily derive that

$$\mathrm{E}\left(\widetilde{\sigma_n^{k1}}(x_0)\right)=0\,.$$

Besides, using the fact that $\mathrm{Cov}(\xi(X_i,\delta_i,s),\xi(X_i,\delta_i,t))=g(\min(s,t))$, change of variable for integraion, integration by parts, and Taylor expansions, we can obtain the covariance of $\left(\widetilde{\sigma_n^{k1}}(x_0),\widetilde{\sigma_n^{m1}}(x_0)\right)$.
Let $K_k(s)=K(s)s^k$, $y=x_0+ht$, and $C_k(p)=\int_{-\infty}^{p}K_k(s)\,ds$,

$$\begin{aligned}
&\mathrm{Cov}\left(\widetilde{\sigma_n^{k1}}(x_0),\widetilde{\sigma_n^{m1}}(x_0)\right)\\
&=\frac{1}{n}h^{k+m}\int\int K_k(s)K_m(t)g(\min(x_0+hs,x_0+ht))\,dL(x_0+hs)\,dL(x_0+ht)\\
&=-\frac{1}{n}h^{k+m}\int K_k(s)\int_{-\infty}^{x_0+hs}L(y)\,d\left[K_m\left(\frac{y-x_0}{h}\right)g(y)\right]\,dL(x_0+hs)\\
&=-\frac{1}{n}h^{k+m}\int L(y)\int_{-\infty}^{\frac{y-x_0}{h}}K_k(s)\,dL(x_0+hs)\\
&\qquad\qquad\left[K_m\left(\frac{y-x_0}{h}\right)dg(y)+g(y)dK_m\left(\frac{y-x_0}{h}\right)\right](1+o(1))\\
&=-\frac{1}{n}h^{k+m+1}\int L(y)l(x_0)C_k\left(\frac{y-x_0}{h}\right)\left[K_m\left(\frac{y-x_0}{h}\right)dg(y)\right.\\
&\qquad\qquad\left.+g(y)dK_m\left(\frac{y-x_0}{h}\right)\right](1+o(1))\\
&=-\frac{1}{n}h^{k+m+1}l(x_0)\left\{\int C_k(t)K_m(t)L(x_0+ht)\,dg(x_0+ht)\right.\\
&\qquad\qquad\left.+\int C_k(t)L(x_0+ht)g(x_0+ht)\,dK_m(t)\right\}(1+o(1))\\
&=\frac{1}{n}h^{k+m+1}l(x_0)\left\{\int C_k(t)K_m(t)g(x_0+ht)\,dL(x_0+ht)\right.\\
&\qquad\qquad\left.+\int K_m(t)K_k(t)L(x_0+ht)g(x_0+ht)\,dt\right\}(1+o(1))\\
&=\frac{1}{n}h^{k+m+1}L(x_0)l(x_0)g(x_0)v_{k+m}(1+o(1))\,.
\end{aligned}$$

Because $\widetilde{\sigma_n^{k1}}(x_0)$ is a sum of i.i.d. random variables, by the Central Limit Theorem, as $n\to\infty$,

$$\sqrt{nh^{-2k-1}}\,\widetilde{\sigma_n^{k1}}(x_0)\xrightarrow{d}\mathcal{N}\Big(0,L(x_0)l(x_0)g(x_0)v_{2k}\Big)\,.$$

□

(c) As for the last two parts of $T_k(x_0)$, since $\sup_x |L_n(x) - L(x)| = O\left((\log n/n)^{1/2}\right)$ a.s., we have the following results.

Lemma 6.3. *Under conditions* (A2) $-$ (A4)*, for* $k = 0, 1, 2$*,*

$$\sup_{0 \le x_0 \le T} \left|\widetilde{\beta_n^{k2}}(x_0)\right| = O\left(h^k \left(\frac{\log n}{n}\right)^{1/2}\right) \quad a.s.$$

$$\mathrm{E}\left(\left|\widetilde{\beta_n^{k2}}(x_0)\right|^r\right) = O\left(\left(\frac{h^k (\log n)^{1/2}}{n^{1/2}}\right)^r\right) \quad \text{for } r = 1, 2\,;$$

and

$$\sup_{0 \le x_0 \le T} \left|\widetilde{\gamma_n^{k2}}(x_0)\right| = O\left(h^k \left(\frac{\log n}{n}\right)^{1/2}\right) \quad a.s.$$

$$\mathrm{E}\left(\left|\widetilde{\gamma_n^{k2}}(x_0)\right|^r\right) = O\left(\left(\frac{h^k (\log n)^{1/2}}{n^{1/2}}\right)^r\right) \quad \text{for } r = 1, 2\,.$$

Then Theorem 3.1 follows from Lemmas 6.1, 6.2, and 6.3.

Proof of Theorem 3.2

Proof. (i) By the asymptotic normality of $T_k(x_0)$, we have

$$\begin{aligned}\sqrt{nh^{-3}}\Big\{ & S_0(x_0)T_1(x_0) - h^2\mu_2 l(x_0)\left[\Lambda(x_0)l'(x_0) + \lambda(x_0)l(x_0)\right] \\ & - \frac{h^4}{6}\{\mu_4 l \left\{3\left[\lambda(x_0)l''(x_0) + \lambda'(x_0)l'(x_0)\right] + \left[\Lambda(x_0)l'''(x_0) + \lambda''(x_0)l(x_0)\right]\right\} \\ & \quad + 3\mu_2^2 l''(x_0)\left[\Lambda(x_0)l'(x_0) + \lambda(x_0)l(x_0)\right]\}\Big\} \\ & \qquad \xrightarrow{d} \mathcal{N}\Big(0, g(x_0)L(x_0)l^3(x_0)v_2\Big)\,.\end{aligned}$$

$$\begin{aligned}\sqrt{nh^{-5}}\Big\{ & S_1(x_0)T_0(x_0) - h^2\mu_2\Lambda(x_0)l(x_0)l'(x_0) \\ & + \frac{h^4}{6}\{\mu_2^2 l'(x_0)\{6\lambda(x_0)l'(x_0) + 3\left[\Lambda(x_0)l\text{''}(x_0) + \lambda'(x_0)l(x_0)\right]\} \\ & \quad - \mu_4\Lambda(x_0)l(x_0)l'''(x_0)\}\Big\} \\ & \qquad \xrightarrow{d} \mathcal{N}\Big(0, g(x_0)L(x_0)l(x_0)l'^2(x_0)\mu_2 v_0\Big)\,.\end{aligned}$$

$$\begin{aligned}\sqrt{nh^{-3}}\Bigg\{ & [S_0(x_0)T_1(x_0) - S_1(x_0)T_0(x_0)] \\ & - \left\{h^2\mu_2 l^2(x_0) + \frac{1}{2}h^4\left[(\mu_4+\mu_2^2)l(x_0)l"(x_0) - 2\mu_2^2 l'^2(x_0)\right]\right\}\lambda(x_0) \\ & - \frac{1}{6}h^4[3(\mu_4-\mu_2^2)\lambda'(x_0)l(x_0)l'(x_0) + \mu_4\lambda''(x_0)l^2(x_0)]\Bigg\} \\ & \xrightarrow{d} \mathcal{N}\Big(0, g(x_0)L(x_0)l^3(x_0)v_2\Big).\end{aligned}$$

$$\begin{aligned}& S_0(x_0)S_2(x_0) - S_1^2(x_0) \\ &= h^2\mu_2 l^2(x_0) + \frac{1}{2}h^4\left[(\mu_4+\mu_2^2)l(x_0)l''(x_0) - 2\mu_2^2 l'^2(x_0)\right] + o_p(h^4).\end{aligned}$$

Hence, we have the result

$$\begin{aligned}\sqrt{nh}\left\{\widehat{\lambda}_{loclin}(x_0) - \lambda(x_0) - \frac{h^2}{6}\frac{3(\mu_4-\mu_2^2)\lambda'(x_0)l'(x_0)/l(x_0) + \mu_4\lambda''(x_0)}{\mu_2}\right\} \\ \xrightarrow{d} \mathcal{N}\left(0, \frac{L(x_0)g(x_0)}{\mu_2^2 l(x_0)}v_2\right).\end{aligned}$$

(ii)

$$\begin{aligned}\sqrt{nh^{-17}}\Bigg\{ & [S_2(x_0)S_3(x_0) - S_1(x_0)S_4(x_0)]\,T_0(x_0) \\ & - \frac{1}{6}h^8(\mu_4^2 - \mu_2\mu_6)\Lambda(x_0)l(x_0)[3l'(x_0)l''(x_0) - l(x_0)l'''(x_0)]\Bigg\} \\ \xrightarrow{d} \mathcal{N}\Bigg(0, & \frac{1}{36}(\mu_4^2-\mu_2\mu_6)^2[3l'(x_0)l''(x_0) - l(x_0)l'''(x_0)]^2 g(x_0)L(x_0)l(x_0)v_0\Bigg).\end{aligned}$$

$$\begin{aligned}\sqrt{nh^{-11}}\Bigg\{ & \left[S_0(x_0)S_4(x_0) - S_2^2(x_0)\right]T_1(x_0) \\ & - h^6\mu_2(\mu_4-\mu_2^2)l^2(x_0)\left[\Lambda(x_0)l'(x_0) + \lambda(x_0)l(x_0)\right] \\ & - \frac{1}{6}h^8\big\{3\mu_2(\mu_6-\mu_2\mu_4)l(x_0)l''(x_0)\left[\Lambda(x_0)l'(x_0) + \lambda(x_0)l(x_0)\right] \\ & \quad + \mu_4(\mu_4-\mu_2^2)l^2(x_0)\left\{3[\lambda(x_0)l''(x_0) + \lambda'(x_0)l'(x_0)]\right. \\ & \quad\quad \left. + [\Lambda(x_0)l'''(x_0) + \lambda''(x_0)l(x_0)]\right\}\big\}\Bigg\} \\ & \xrightarrow{d} \mathcal{N}\Big(0, (\mu_4-\mu_2^2)^2 g(x_0)L(x_0)l^5(x_0)v_2\Big).\end{aligned}$$

$$
\begin{aligned}
\sqrt{nh^{-13}}\Big\{ & [S_1(x_0)S_2(x_0) - S_0(x_0)S_3(x_0)]\, T_2(x_0) \\
& - h^6\mu_2(\mu_2^2 - \mu_4)\Lambda(x_0)l^2(x_0)l'(x_0) \\
& - \frac{1}{6}h^8\big\{\mu_4(\mu_2^2 - \mu_4)l(x_0)l'(x_0)\big\{6\lambda(x_0)l'(x_0) \\
& \quad + 3\left[\Lambda(x_0)l''(x_0) + \lambda'(x_0)l(x_0)\right]\big\} \\
& \qquad + \mu_2(\mu_2\mu_4 - \mu_6)\Lambda(x_0)l^2(x_0)l'''(x_0)\big\}\Big\} \\
& \qquad\qquad \xrightarrow{d} \mathcal{N}\Big(0, (\mu_4 - \mu_2^2)^2 g(x_0)L(x_0)l^3(x_0)l'^2(x_0)v_4\Big).
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{nh^{-11}}\Big\{ & [S_2(x_0)S_3(x_0) - S_1(x_0)S_4(x_0)]\, T_0(x_0) \\
& + \left[S_0(x_0)S_4(x_0) - S_2^2(x_0)\right] T_1(x_0) \\
& + [S_1(x_0)S_2(x_0) - S_0(x_0)S_3(x_0)]\, T_2(x_0) \\
& - \big\{h^6\mu_2(\mu_4 - \mu_2^2)l^3(x_0) - \frac{1}{2}h^8\left\{(\mu_2\mu_6 + \mu_4^2)l^2(x_0)l''(x_0)\right. \\
& \quad \left. + 2\mu_4(\mu_2^2 - \mu_4)l(x_0)l'^2(x_0) - 2\mu_2^2\mu_4 l^2(x_0)l''(x_0)\right\}\big\}\lambda(x_0) \\
& \qquad + \frac{1}{6}h^8\mu_4(\mu_4 - \mu_2^2)\lambda''(x_0)l^3(x_0)\Big\} \\
& \qquad\qquad \xrightarrow{d} \mathcal{N}\Big(0, (\mu_4 - \mu_2^2)^2 g(x_0)L(x_0)l^5(x_0)v_2\Big).
\end{aligned}
$$

$$
\begin{aligned}
\Delta(x_0) = {} & h^6\mu_2(\mu_4 - \mu_2^2)l^3(x_0) + \frac{1}{2}h^8\left\{(\mu_2\mu_6 + \mu_4^2)l^2(x_0)l''(x_0)\right. \\
& \left. + 2\mu_4(\mu_2^2 - \mu_4)l(x_0)l'^2(x_0) - 2\mu_2^2\mu_4 l^2(x_0)l''(x_0)\right\} + o_p(h^8).
\end{aligned}
$$

Hence, we have the result

$$
\sqrt{nh}\left\{\widehat{\lambda}_{locqua}(x_0) - \lambda(x_0) - \frac{h^2}{6}\frac{\mu_4}{\mu_2}\lambda''(x_0)\right\} \xrightarrow{d} \mathcal{N}\left(0, \frac{L(x_0)g(x_0)}{\mu_2^2 l(x_0)}v_2\right).
$$

□

References

Cheng, M.-Y., Peng, L., and Sun, S. (2006). Variance reduction in hazard function estimation. *International Journal of Statistics and Systems*, **1**, 87–110.

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

Jiang, J. and Doksum, K. (2003). On local polynomial estimation of hazard rates and their derivatives under random censoring. *In Mathematical Statistics and Applications: Festschrift for Constance van Eeden. IMS*, 463–481.

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

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.

Loader, C. (1999). *Local Regression and Likelihood*. Springer.

Müller, H. G. and Wang, J. L. (1990). Locally adaptive hazard smoothing. *Probability Theory and Related 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.

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

[Received March 2009; accepted April 2009.]