Variance Reduction in Hazard Function Estimation

Ming-Yen Cheng[1], **Liang Peng**[2] **and Shan Sun**[3]

December, 2005

Abstract

In this paper we propose to estimate the hazard function based on local smoothing techniques for both i.i.d and censoring data. Such estimators are known to have no boundary effects while the estimators based on kernel function have the boundary effect, as pointed out by Müller and Wang (1990). We derive the asymptotic normalities of the local smooth estimators and compare with the kernel smooth estimators. It turns out that our local smooth estimators with optimal bandwidths produce smaller biases than that of the kernel smooth estimators. However, such estimators have large variances than that of the kernel smooth estimators. To overcome this problem, we apply the variance reduction technique in Cheng, Peng and Wu (2005) to our estimators. The resulted estimators have the same asymptotic biases as the local smooth estimators and smaller asymptotic variances than the kernel estimators.

Keywords. Hazard function, kernel smooth estimation, local polynomial estimation, variance reduction.

1 Introduction

Hazard function based on i.i.d. or censored data is important. It provides useful information in reliability theory and survival analysis, as well as in the fields as diverse as engineering,

[1]Department of Mathematics, National Taiwan University, Taipei 106, Taiwan. Email: cheng@math.ntu.edu.tw
[2]School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332-0160, USA. Email: peng@math.gatech.edu
[3]Department of Mathematics and Statistics, Texas Tech University, Lubbock, Texas 79409 - 1042, USA. Email: ssun@math.ttu.edu

medical statistics and geophysics. A variety of inferential procedures have been proposed to estimate the hazard function nonparametrically. Estimators of hazard function based on kernel smooth estimation have been studied extensively in the literature. For related investigations in this direction we refer to Watson and Leadbetter (1964), Murphy (1965), Rice and Rosenblatt (1976), Singpurwalla and Wong (1983) and Patil (1997), under the i.i.d case. Under the censoring case, see the discussions in Tanner and Wong (1983), Tanner (1983), Schäfer (1985), Liu and Ryzin (1985), Diehl and Stute (1988), Lo, Mack and Wang (1989), Müller and Wang (1990), Patil (1993), Wang (1999). It was pointed out that the drawback of using the kernel smooth estimators are known to have the boundary effect (see by Müller and Wang (1990)).

Recently, Jiang and Doksum (2003) propose a type of local polynomial estimation for hazard rates and their derivatives via smoothing a Dirac derivative of the Nelson-Aalen estimator. The result is the same as the kernel estimator using the equivalent kernel of local polynomial regression, see e.g. Fan and Gijbels (1996), and hence is free from boundary effects. In this paper we propose another type of smooth estimation for the hazard function based on local polynomial techniques, which is more intuitive than Jiang and Doksum (2003) and may be argued to have no boundary effect as well. We show that our local smooth estimators have smaller asymptotic biases, but larger asymptotic variances, than the kernel smooth estimators under the case with or without censoring. Proofs of these results are nontrivial. To reduce the variances of our local smooth estimators, we apply the variance reduction technique introduced by Cheng, Peng and Wu (2005). Hence, our variance reduced local smooth estimators are better than both the kernel smooth estimators and that in Jiang and Doksum (2003) in terms of either optimal asymptotic mean squared error or asymptotic bias and variance with the same bandwidth. A numerical study demonstrates that these advantageous asymptotic properties are also apparent in finite sample sizes.

We organize this paper as follows. In Section 2, we establish the weak convergence of the local smooth estimators for both i.i.d and censoring cases. In Section 3, we provide some comparisons between our local smooth estimators and the kernel smooth estimators. In Section 4, we propose variance reduced local smooth estimators and compare them with the kernel smooth estimators. A simulation study is given in Section 5. All proofs are deferred till Section 6.

2 Local smooth estimation

Throughout this paper we assume that

A1) $k(x)$ is a symmetric density function with support [-1, 1];

A2) $f''(x)$ exists and is continuous, where f is defined below;

A3) $h = h(n) > 0$, $h \to 0$ and $\sqrt{nh}h^2 \to b \in [0, \infty)$ as $n \to \infty$.

2.1. The case without censoring. Let $X_1, \cdots, X_n$ be independent and identically distributed survival times with distribution function $F(x)$ and density function $f(x)$. Our aim is to estimate the hazard function $\lambda(x) = \frac{f(x)}{1-F(x)}$ using local smooth techniques. We apply local smoothing techniques, see for example Fan and Gijbels (1996), to estimate the derivative of $\Lambda(x) = -\log(1 - F(x))$, i.e., $\lambda(x)$ as follows.

Let $F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$ be the empirical distribution of the sample $\{X_1, X_2, ..., X_n\}$ and define $\Lambda_n(x) = -\log(1 - F_n(x))$. Observe the following regression model:

$$\Lambda_n(X_i) = \Lambda(X_i) + \text{error}, i = 1, \cdots, n,$$

and let $(\hat{a}, \hat{b}, \hat{c})$ be the value of (a, b, c) that minimizes the following kernel weighted squared errors:

$$\sum_{j=1}^{n} \{\Lambda_n(X_j) - a - b(X_j - x) - c(X_j - x)^2\}^2 k(\frac{x - X_j}{h}). \tag{2.1}$$

Then our new local smooth estimator for $\lambda(x)$ is defined as $\hat{\lambda}_n(x) = \hat{b}$ and has the following explicit expression

$$\hat{\lambda}_n(x) = \frac{\sum_{j=1}^{n} \Lambda_n(X_j) k(\frac{x - X_j}{h})[\Delta_{n,1}(x) + (X_j - x)\Delta_{n,2}(x) + (X_j - x)^2 \Delta_{n,3}(x)]}{\Delta_n(x)},$$

where $s_{n,l}(x) = \sum_{j=1}^{n}(x - X_j)^l k(\frac{x - X_j}{h})$, $l = 0, 1, 2, 3, 4$, and

$$\begin{cases} \Delta_n(x) & = s_{n,1}^3(x)s_{n,4}(x) + s_{n,1}(x)s_{n,2}^3(x) + s_{n,0}(x)s_{n,1}(x)s_{n,3}^2(x) \\ & \quad -2s_{n,1}^2(x)s_{n,2}(x)s_{n,3}(x) - s_{n,0}(x)s_{n,1}(x)s_{n,2}(x)s_{n,4}(x) \\ \Delta_{n,1}(x) & = s_{n,1}(x)s_{n,2}(x)s_{n,3}(x) - s_{n,1}^2(x)s_{n,4}(x) \\ \Delta_{n,2}(x) & = s_{n,1}(x)s_{n,2}^2(x) - s_{n,0}(x)s_{n,1}(x)s_{n,4}(x) \\ \Delta_{n,3}(x) & = s_{n,1}^2(x)s_{n,2}(x) - s_{n,0}(x)s_{n,1}(x)s_{n,3}(x). \end{cases} \tag{2.2}$$

Define

$$c_1 = \int_{-1}^{1} s^2 k(s)\, ds, \quad c_2 = \int_{-1}^{1} s^4 k(s)\, ds, \quad c_3 = 2\int_{-1}^{1} \Big\{ \int_{-1}^{t} k(s)k(t)s^2 t\, ds \Big\}\, dt. \tag{2.3}$$

The following theorem provides us with the weak convergence of the local smooth estimator $\hat{\lambda}_n(x)$.

Theorem 1. *Under regularity conditions* **A1)** $-$ **A3)**, *we have for* $1 - F(x) > 0$

$$\sqrt{nh}\{\hat{\lambda}_n(x) - \lambda(x)\} \xrightarrow{d} N\Big(\frac{b\lambda''(x)c_2}{6c_1}, \frac{f(x)c_3}{[1 - F(x)]^2 c_1^2}\Big)$$

as $n \to \infty$, *where* b *is defined in condition A3 and* c_1, c_2 *and* c_3 *are defined in (2.3).*

2.2. The case with censoring. Let $X_1, \cdots, X_n$ be independent and identically distributed random variables with distribution function $F(x)$ and density function $f(x)$, and $Y_1, \cdots, Y_n$ be independent and identically distributed random variables with distribution function $G(y)$ and density function $g(y)$. Suppose $X_i's$ and $Y_j's$ are independent and our observations are $Z_i = \min(X_i, Y_i)$ with censoring indicators $\delta_i = I(X_i \le Y_i)$ for $i = 1, \cdots, n$. Thus $\delta_i = 1$ indicates the survival time X_i for the ith individual is observed while $\delta_i = 0$ indicates X_i is not observed but it is known to be greater than Y_i. Our aim is to estimate the hazard function $\lambda(x) = \frac{f(x)}{1 - F(x)}$, which is of importance in many lifetime studies.

In estimating the distribution function F, a popular nonparametric estimator F_n^* based on the right censored data $\{Z_i, \delta_i\}, i = 1, ..., n$, is the well-known Kaplan-Meier (Kaplan and

Meier, 1958) estimator given by

$$F_n^*(x) = \begin{cases} 1 - \prod_{j=1}^{n} [\frac{N(Z_j)}{1+N(Z_j)}]^{I(Z_j \le x, \delta_j = 1)} & \text{if} \quad x < \max(Z_1, \cdots, Z_n) \\ 1 & \text{elsewhere,} \end{cases}$$

where $N(u) = \sum_{j=1}^{n} I(Z_j > u)$. The large sample properties of the product-limit estimator $F_n^*(x)$ have drawn much attention in the literature; see Chen and Lo (1997) and references cited therein.

Define $\Lambda_n^*(x) = -\log(1 - F_n^*(x))$ and $\Lambda(x) = -\log(1 - F(x))$. Observe the following regression model:

$$\Lambda_n^*(Z_i) = \Lambda(Z_i) + \text{ error }, i = 1, \cdots, n.$$

For those $i's$ such that $\delta_i = 1$, i.e., $Z_i = X_i$, we apply local smoothing techniques to estimate the derivative of $\Lambda(x)$, i.e., $\lambda(x)$. That is, let $(\hat{a}, \hat{b}, \hat{c})$ be the value of (a, b, c) that minimizes the following kernel weighted squared errors:

$$\sum_{j=1}^{n} \delta_j \{\Lambda_n^*(Z_j) - a - b(Z_j - x) - c(Z_j - x)^2\}^2 k(\frac{x - Z_j}{h}). \tag{2.4}$$

Then our local smooth estimator for $\lambda(x)$ is defined as $\hat{\lambda}_n^*(x) = \hat{b}$ and has the following explicit expression

$$\hat{\lambda}_n^*(x) = \frac{\sum_{j=1}^{n} \Lambda_n^*(Z_j) k(\frac{x - Z_j}{h})[\Delta_{n,1}(x) + (Z_j - x)\Delta_{n,2}(x) + (Z_j - x)^2 \Delta_{n,3}(x)]}{\Delta_n(x)}, \tag{2.5}$$

where $s_{n,l}(x) = \sum_{j=1}^{n} \delta_j (x - Z_j)^l k(\frac{x - Z_j}{h})$, $l = 0, 1, 2, 3, 4$, and $\Delta_n(x)$, $\Delta_{n,1}(x)$, $\Delta_{n,2}(x)$ and $\Delta_{n,3}(x)$ are defined as in (2.2). The following theorem provides us the week convergence of the local smooth estimator $\hat{\lambda}_n^*(x)$.

Theorem 2. *Under regularity conditions* **A1) - A3)** *and $g(y)$ is continuous, we have for* $1 - F(x) > 0$

$$\sqrt{nh}\{\hat{\lambda}_n^*(x) - \lambda(x)\} \xrightarrow{d} N\Big(\frac{b\lambda''(x)c_2}{6c_1}, \frac{f(x)c_3}{[1 - F(x)]^2[1 - G(x)]c_1^2}\Big)$$

as $n \to \infty$, where b is defined in condition A3 and c_1, c_2 and c_3 are given by (2.3).

Remark 1. *In the local polynomial fittings (2.1) and (2.4), Λ_n and Λ_n^* can be replaced by the Nelson-Aalen estimator of the cumulative hazard. The asymptotic results remain unchanged.*

3 Comparisons between kernel smooth estimators and local smooth estimators

In this section, we study some asymptotic properties of the kernel smooth estimators and local smooth estimators for both i.i.d and censored cases.

3.1. The case without censoring. Under the case of no censoring, the kernel smooth estimator studied by Singpurwalla and Wong (1983) is defined as

$$\bar{\lambda}_n(x) = \frac{1}{h}\sum_{j=1}^{n} k(\frac{X_{(j)} - x}{h})/(n - j + 1),$$

where $X_{(1)} \leq \cdots \leq X_{(n)}$ denote the order statistics of $X_1, \cdots, X_n$. Under the regularity conditions **A1) - A3)**, Singpurwalla and Wong (1983) proved that for $1 - F(x) > 0$

$$\sqrt{nh}(\bar{\lambda}_n(x) - \lambda(x)) \xrightarrow{d} N\Big(\frac{b\lambda''(x)c_1}{2}, \frac{f(x)c_4}{[1 - F(x)]^2}\Big)$$

as $n \to \infty$, where c_1 is given by (2.3), b satisfies condition **A3** and

$$c_4 = \int_{-1}^{1} k^2(x)dx.$$

Hence, by minimizing the asymptotic mean squared error, we obtain that the local optimal bandwidth for $\bar{\lambda}_n(x)$ is

$$\bar{h}_{opt} = n^{-1/5}\Big\{\frac{f(x)c_4}{[1 - F(x)]^2[\lambda''(x)]^2c_1^2}\Big\}^{1/5}. \tag{3.1}$$

Thus the optimal asymptotic mean squared error of $\bar{\lambda}_n(x)$ is given by

$$amse(\bar{\lambda}_n(x), \bar{h}_{opt}) = n^{-4/5}\Big\{\frac{f(x)}{[1 - F(x)]^2}\Big\}^{4/5}\{\lambda''(x)\}^{2/5}\frac{5c_1^{2/5}c_4^{4/5}}{4}. \tag{3.2}$$

On the other hand, it follows from Theorem 1 that, by minimizing the asymptotic mean squared error (amse), the local optimal bandwidth for our local smooth estimator $\hat{\lambda}_n(x)$ is

$$\hat{h}_{opt} = n^{-1/5}\Big\{\frac{f(x)9c_3}{[1 - F(x)]^2[\lambda''(x)]^2c_2^2}\Big\}^{1/5}. \tag{3.3}$$

Thus the optimal asymptotic mean squared error for $\hat{\lambda}_n(x)$ is given by

$$amse(\hat{\lambda}_n(x), \hat{h}_{opt}) = n^{-4/5}\Big\{\frac{f(x)}{[1-F(x)]^2}\Big\}^{4/5}\{\lambda''(x)\}^{2/5}\frac{5c_3^{4/5}c_2^{2/5}}{4\cdot 3^{2/5}c_1^2}. \tag{3.4}$$

3.2. The case with censoring. For the case of censoring, a kernel smooth estimator for $\lambda(x)$ was proposed by Tanner and Wong (1983) as

$$\bar{\lambda}_n^*(x) = \frac{1}{h}\sum_{j=1}^{n}(n-j-1)^{-1}\delta_{(j)}k(\frac{x-Z_{(j)}}{h}),$$

where $Z_{(1)} \le \cdots \le Z_{(n)}$ denote the order statistics of $Z_1, \cdots, Z_n$ and $\delta_{(1)}, \cdots, \delta_{(n)}$ denote the corresponding censoring indicators. Under the same regularity conditions as in Theorem 2, Müller and Wang (1990) showed that for $1-F(x) > 0$

$$\sqrt{nh}(\bar{\lambda}_n^*(x) - \lambda(x)) \xrightarrow{d} N\left(\frac{b\lambda''(x)c_1}{2}, \frac{f(x)c_4}{[1-F(x)]^2[1-G(x)]}\right)$$

as $n \to \infty$. Hence, by minimizing the asymptotic mean squared error of $\bar{\lambda}_n^*(x)$, the local optimal bandwidth for kernel smooth estimator $\bar{\lambda}_n^*(x)$ is given by

$$\bar{h}_{opt}^* = n^{-1/5}\Big\{\frac{f(x)c_4}{[1-F(x)]^2[1-G(x)][\lambda''(x)]^2c_1^2}\Big\}^{1/5}. \tag{3.5}$$

Thus, the optimal asymptotic mean squared error for $\bar{\lambda}_n^*(x)$ is

$$amse(\bar{\lambda}_n^*(x), \bar{h}_{opt}^*) = n^{-4/5}\Big\{\frac{f(x)}{[1-F(x)]^2[1-G(x)]}\Big\}^{4/5}\{\lambda''(x)\}^{2/5}\frac{5c_1^{2/5}c_4^{4/5}}{4}. \tag{3.6}$$

On the other hand, it follows from Theorem 2 that, by minimizing the asymptotic mean squared error (amse), the local optimal bandwidth for local smooth estimator $\hat{\lambda}_n^*(x)$ is

$$\hat{h}_{opt}^* = n^{-\frac{1}{5}}\Big\{\frac{f(x)9c_3}{[1-F(x)]^2[1-G(x)][\lambda''(x)]^2c_2^2}\Big\}^{\frac{1}{5}}. \tag{3.7}$$

Therefore, the optimal asymptotic mean squared error for $\hat{\lambda}_n^*(x)$ is given by

$$amse(\hat{\lambda}_n^*(x), \hat{h}_{opt}^*) = n^{-\frac{4}{5}}\Big\{\frac{f(x)}{[1-F(x)]^2[1-G(x)]}\Big\}^{\frac{4}{5}}\{\lambda''(x)\}^{\frac{2}{5}}\frac{5c_3^{4/5}c_2^{2/5}}{4\cdot 3^{2/5}c_1^2}. \tag{3.8}$$

3.3. Comparisons. For the purpose of comparison, we compute the values of $c_1, \cdots, c_4$ and the contant factors in the asymptotically optimal bandwidth and amse expressions (3.1)–(3.8) for four commonly used kernels, They are the Epanechnikov, Biweight, Triangular and Uniform kernels. The results are shown in Table 1.

Table 1: *Values of c_1, c_2, c_3, c_4 for some commonly used kernels.*

Kernel	$\frac{c_2^2}{9c_1^2}$	c_1^2	$\frac{c_3}{c_1^2}$	c_4	$\frac{c_3^{4/5}c_2^{2/5}}{3^{2/5}c_1^2}$	$c_1^{2/5}c_4^{4/5}$	$\left(\frac{9c_3}{c_2^2}\right)^{1/5}$	$\left(\frac{c_4}{c_1^2}\right)^{1/5}$
Epanechnikov $k(x) = \frac{3}{4}(1 - x^2)I(\lvert x\rvert \leq 1)$	$\frac{1}{49}$	$\frac{1}{25}$	$\frac{5}{7}$	$\frac{3}{5}$	0.350799	0.3490865	2.036168	1.718772
Biweight $k(x) = \frac{15}{16}(1 - x^2)^2 I(\lvert x\rvert \leq 1)$	$\frac{1}{81}$	$\frac{1}{49}$	0.81585	$\frac{5}{7}$	0.3528509	0.350799	2.312166	2.036168
Triangular $k(x) = (1 - \lvert x\rvert)I(\lvert x\rvert \leq 1)$	$\frac{9}{100}$	$\frac{1}{36}$	$\frac{26}{35}$	$\frac{2}{3}$	0.3521273	0.3530746	2.109626	1.888175
Uniform $k(x) = \frac{1}{2}I(\lvert x\rvert \leq 1)$	$\frac{1}{25}$	$\frac{1}{9}$	$\frac{3}{5}$	$\frac{1}{2}$	0.3490865	0.3701072	1.718772	1.350960

First, let us look at the optimal *amse* for kernel smooth estimators and our local smooth estimators. From (3.2), (3.4), (3.6) and (3.8), we notice that for both i.i.d case and censored case, the difference in *amse* between these estimators are mainly the terms $\frac{c_3^{4/5}c_2^{2/5}}{3^{2/5}c_1^2}$ and $c_1^{2/5}c_4^{4/5}$. Table 1 shows that our local smooth estimators ($\hat{\lambda}_n$ and $\hat{\lambda}_n^*$) have smaller bias terms than that of the kernel estimators ($\bar{\lambda}_n$ and $\bar{\lambda}_n^*$), but with large variances in general, for the four commonly used kernel functions. It is also interesting to see that the optimal mean squared error is the same for local smooth estimators with uniform kernel and for the kernel smooth estimators with Epanechnikov kernel. In fact, this optimal mean squared error is the smallest among the four kernels for both local and kernel smooth estimators.

Next, let us compare the optimal bandwidths for both local ($\hat{\lambda}_n$ and $\hat{\lambda}_n^*$) and kernel smooth estimators ($\bar{\lambda}_n$ and $\bar{\lambda}_n^*$). Observe equations (3.1), (3.3), (3.5) and (3.7), we see that, for both i.i.d. case and censoring case, the difference in optimal bandwidths for kernel and local smooth estimators are based on terms $\left(\frac{c_4}{c_1^2}\right)^{1/5}$ and $\left(\frac{9c_3}{c_2^2}\right)^{1/5}$. From Table 1, we see that the optimal bandwidths for our local smooth estimators are larger than those for the kernel smooth estimators. So, in practice, one may prefer local smooth estimators to kernel smooth estimators since the larger optimal bandwidth will allow more data points in the local model.

4 Variance reduced local smooth estimation

Note that our local smooth estimator has a smaller asymptotic bias, but a larger asymptotic variance, than the kernel smooth estimator under the case with or without censoring. After tedious calculations, we also notice that our local smooth estimator has the same optimal asymptotic mean squared error as the local linear smooth estimator in Jiang and Doksum (2003). However, we are able to reduce the asymptotic variance and retain the asymptotic bias of our local smooth estimator by employing the variance reduction technique in Cheng, Peng and Wu (2005); see below for details.

4.1. The case without censoring. We consider the following variance reduced local smooth estimators

$$\tilde{\lambda}_n(x) = \frac{1-\sqrt{2}}{4}\hat{\lambda}_n(x-(\sqrt{1/2}+1)\delta h) + \frac{1}{2}\hat{\lambda}_n(x-\sqrt{1/2}\delta h) + \frac{1+\sqrt{2}}{4}\hat{\lambda}_n(x-(\sqrt{1/2}-1)\delta h)$$

where $\delta > 0$. This estimator is a linear combination of the three values $\hat{\lambda}_n(x-(\sqrt{1/2}+1)$, $\hat{\lambda}_n(x-\sqrt{1/2}\delta h)$ and $\hat{\lambda}_n(x-(\sqrt{1/2}-1)\delta h)$, and it is parallel to the form of the variance reduced local linear regression estimator of Cheng, Peng and Wu (2005). The principle of Cheng, Peng and Wu (2005) is to find the maximal relative variance reduction among all points in an interpolation interval of length $2\delta h$. In the current hazard estimation context, the covariance structure of the local smooth estimator at different locations is much more complicated than in the regression setting. For simplicity reasons we take $\tilde{\lambda}_n(x)$ the specified form. This may not achieve the most variance reduction. Nevertheless, $\tilde{\lambda}_n(x)$ admits a very simple form and it is shown that $\tilde{\lambda}_n(x)$ enjoys superior performance in both asymptotic and finite sample cases.

To analyze asymptotic properties of the new estimator, define

$$c_5(a,b) = \int_{-1}^{1}\int_{-1}^{t-a+b} k(s)k(t)(s-b)st\,dsdt + \int_{-1}^{1}\int_{-1}^{t+a-b} k(s)k(t)(s-a)st\,dsdt \tag{4.1}$$

and

$$\begin{aligned} c_6(\delta) \quad = \quad & \tfrac{5}{8}c_3 + \tfrac{1-\sqrt{2}}{4}c_5((\sqrt{1/2}+1)\delta, \sqrt{1/2}\delta) - \tfrac{1}{8}c_5((\sqrt{1/2}+1)\delta, (\sqrt{1/2}-1)\delta) \\ & + \tfrac{1+\sqrt{2}}{4}c_5(\sqrt{1/2}\delta, (\sqrt{1/2}-1)\delta). \end{aligned} \tag{4.2}$$

First we derive the asymptotic normality for our variance reduced local smooth estimator as follows.

Theorem 3. *Under regularity conditions* **A1) - A3)**, *we have for* $1 - F(x) > 0$ *and* $\delta > 0$

$$\sqrt{nh}\{\tilde{\lambda}_n(x) - \lambda(x)\} \xrightarrow{d} N\Big(\frac{b\lambda''(x)c_2}{6c_1}, \frac{f(x)c_6(\delta)}{[1-F(x)]^2c_1^2}\Big)$$

as $n \to \infty$.

Second, we shall compare our variance reduced local smooth estimator $\tilde{\lambda}_n(x)$ with the kernel smooth estimator $\bar{\lambda}_n(x)$ defined in Section 3.1. Since the local smooth estimator $\hat{\lambda}_n(x)$ with the uniform kernel and $\bar{\lambda}_n(x)$ with the Epanechnikov kernel have the same smallest optimal amse among the four different kernels considered in Section 3.3, we only compare between the variance reduced local smooth estimator with the uniform kernel and the kernel smooth estimator with the Epanechnikov kernel. In this case, we have for $0 \le b - a \le 2$

$$\begin{aligned} c_5(a,b) &= \tfrac{1}{4}\textstyle\int_{-1-a+b}^{1}\{\int_{-1}^{t+a-b}(s-a)st\,ds\}\,dt + \tfrac{1}{4}\int_{-1-a+b}^{1}\{\int_{t+a-b}^{1}(t-b)st\,ds\}\,dt \\ &\quad +\tfrac{1}{4}\textstyle\int_{-1}^{-1-a+b}\{\int_{-1}^{1}(t-b)st\,ds\}\,dt \\ &= \tfrac{1}{4}\textstyle\int_{-1-a+b}^{1} t\{\frac{(t+a-b)^3+1}{3} - \frac{a(t+a-b)^2-a}{2}\}\,dt + \tfrac{1}{4}\int_{-1-a+b}^{1} t(t-b)\frac{1-(t+a-b)^2}{2}\,dt \\ &= \tfrac{1}{4}\textstyle\int_{-1-a+b}^{1}\{-\frac{t^4}{6} + t^3\frac{b-a}{2} + t^2(-\frac{(b-a)^2}{2}+\frac{1}{2}) + t(\frac{(b-a)^3}{b} - \frac{b-a}{2} + \frac{1}{3})\}\,dt \\ &= \tfrac{1}{15} - \tfrac{(b-a)^2}{12} + \tfrac{(b-a)^3}{24} - \tfrac{(b-a)^5}{480}, \end{aligned}$$

and for $b - a > 2$, $c_5(a,b) = 0$. Therefore,

$$c_6(\delta) = \begin{cases} 1/15 - \delta^3/48 + 7\delta^5/960 & \text{if } 0 \le \delta \le 1 \\ 3/40 - \delta^2/24 + \delta^3/48 - \delta^5/960 & \text{if } 1 < \delta \le 2 \\ 1/24 & \text{if } \delta > 2, \end{cases}$$

i.e.,

$$c_6(\delta)/c_1^2 = \begin{cases} 3/5 - 3\delta^3/16 + 21\delta^5/320 & \text{if } 0 \le \delta \le 1 \\ 27/40 - 3\delta^2/8 + 3\delta^3/16 - 3\delta^5/320 & \text{if } 1 < \delta \le 2 \\ 3/8 & \text{if } \delta > 2. \end{cases}$$

Notice that $\tilde{\lambda}_n(x)$ with $\delta = 0$ reduces to the original estimator $\hat{\lambda}_n(x)$. By checking that $\frac{d}{d\delta}c_6(\delta) \le 0$, $\delta > 0$, we see that $c_6(\delta)/c_1^2 < \frac{3}{5}$ for any $\delta > 0$, i.e., $\tilde{\lambda}_n(x)$ with the uniform kernel has a smaller asymptotic variance than $\bar{\lambda}_n(x)$ with the Epanechnikov kernel. Because both estimators have the same asymptotic bias, we conclude that the variance reduced local smooth estimator $\tilde{\lambda}_n(x)$ with the uniform kernel is better than the kernel smooth estimator $\bar{\lambda}(x)$ with the Epanechnikov kernel in terms of optimal amse or amse with the same bandwidth.

Cheng, Peng and Wu (2005) discussed in detail the choice of the parameter δ. Larger values of δ are preferred so that more variance reductions are achieved if the hazard function is smooth. Otherwise, if the curve has sharp feature, second order bias may appear and play a role. In that case, smaller values of δ would still provide reasonable amount of variance reductions.

4.2. The case with censoring.

The variance reduced local smooth estimators in this case is defined as

$$\tilde{\lambda}_n^*(x) = \frac{1-\sqrt{2}}{4}\hat{\lambda}_n^*(x-(\sqrt{1/2}+1)\delta h) + \frac{1}{2}\hat{\lambda}_n^*(x-\sqrt{1/2}\delta h) + \frac{1+\sqrt{2}}{4}\hat{\lambda}_n^*(x-(\sqrt{1/2}-1)\delta h),$$

where $\delta > 0$. The asymptotic normality of this variance reduced local smooth estimator is given below. The comparison between $\tilde{\lambda}_n(x)$ and $\bar{\lambda}_n(x)$ is similar to the i.i.d case in Section 4.1, hence is omitted here.

Theorem 4. *Under regularity conditions* **A1) - A3)** *and that $g(y)$ is continuous, we have for $1-F(x)>0$ and any $\delta > 0$*

$$\sqrt{nh}\{\tilde{\lambda}_n^*(x) - \lambda(x)\} \xrightarrow{d} N\Big(\frac{b\lambda''(x)c_2}{6c_1}, \frac{f(x)c_6(\delta)}{[1-F(x)]^2[1-G(x)]c_1^2}\Big)$$

as $n \to \infty$, where c_6 is defined as in (4.2)

5 Simulation study

A Monte Carlo study was conducted to demonstrate the advantage of our variance reduced local smooth estimator $\tilde{\lambda}_n(x)$ over the kernel smooth estimator $\bar{\lambda}_n(x)$, under the i.i.d setup. The uniform kernel and Epanechnikov kernel were employed for $\tilde{\lambda}_n(x)$ and $\bar{\lambda}_n(x)$, respectively. Moreover, value of δ in the definition of $\tilde{\lambda}_n(x)$ was taken as one.

We generated 1000 pseudo-random samples of size $n = 100$ from Weibull distribution $F(x) = 1 - \exp(-x^\alpha)$, $x \geq 0$. We took $\alpha = 4$ and compute $\tilde{\lambda}_n(x)$ and $\bar{\lambda}_n(x)$ at point x such that $F(x) = 0.5$ for $h = \frac{\bar{h}_{opt}}{2} + \frac{j}{20}\bar{h}_{opt}, j = 0, 1, \cdots, 19$, where $\bar{h}_{opt}$ is defined in (3.1). In Figure 1, we plot the mean squared errors of $\bar{\lambda}_n(x)$ and $\tilde{\lambda}_n(x)$ against different h. This figure clearly shows that $\tilde{\lambda}_n(x)$ has a substantially smaller mean squared error than $\bar{\lambda}_n(x)$. This confirms the asymptotic results.

6 Proofs

Proof of Theorem 1. Let $U_j = F(X_j), j = 1, \cdots, n$, $G_n(u) = \frac{1}{n}\sum_{i=1}^{n} I(U_i \leq u)$ and $\alpha_n(u) = \sqrt{n}(G_n(u) - u)$. Then, using the result of Komlós, Major and Tusnády (1975), there exists a sequence of Brownian bridges $B_n(u)$, $0 \leq u \leq 1$, $n = 1, 2, \cdots$, such that

$$\sup_{0\leq u\leq 1} |\alpha_n(u) - B_n(u)| = O_p(n^{-1/2}\log n). \tag{6.1}$$

Note that

$$\begin{cases} n^{-1}h^{-1}s_{n,0}(x) = f(x) + O_p(h^2) \\ n^{-1}h^{-3}s_{n,1}(x) = -f'(x)c_1 + O_p(h) \\ n^{-1}h^{-3}s_{n,2}(x) = f(x)c_1 + O_p(h^2) \\ n^{-1}h^{-5}s_{n,3}(x) = -f'(x)c_2 + O_p(h) \\ n^{-1}h^{-5}s_{n,4}(x) = f(x)c_2 + O_p(h^2). \end{cases} \tag{6.2}$$

Using (6.2) we obtain that

$$\begin{cases} \Delta_{n,1}(x) = o_p(n^3h^{11}) \\ \Delta_{n,2}(x) = n^3h^9f^2(x)f'(x)c_1(c_2 - c_1^2) + O_p(n^3h^{10}) \\ \Delta_{n,3}(x) = n^3h^9f(x)[f'(x)]^2c_1(c_1^2 - c_2) + O_p(n^3h^{10}) \\ \Delta_n(x) = n^4h^{12}f^3(x)f'(x)c_1^2(c_2 - c_1^2) + O_p(n^4h^{13}). \end{cases} \tag{6.3}$$

Since

$$
\begin{aligned}
& \{\hat{\lambda}_n(x) - \lambda(x)\}\Delta_n(x) \\
= \; & \sum_{j=1}^{n} \{\Lambda_n(X_j) - \Lambda(X_j)\} k(\frac{x - X_j}{h}) \{\Delta_{n,1}(x) + (X_j - x)\Delta_{n,2}(x) + (X_j - x)^2 \Delta_{n,3}(x)\} \\
& + \sum_{j=1}^{n} \{\Lambda(X_j) - \Lambda(x) - \lambda(x)(X_j - x) - \frac{1}{2}\lambda'(x)(X_j - x)^2\} k(\frac{x - X_j}{h}) \\
& \quad \times \{\Delta_{n,1}(x) + (X_j - x)\Delta_{n,2}(x) + (X_j - x)^2 \Delta_{n,3}(x)\} \\
= \; & I + II,
\end{aligned}
$$

it is easy to see that Theorem 1 holds if we show that

$$\sqrt{nh} n^{-4} h^{-12} I \xrightarrow{d} N\left(0, \frac{f^7(x)[f'(x)]^2 c_1^2 (c_2 - c_1^2)^2 c_3}{[1 - F(x)]^2}\right) \tag{6.4}$$

and

$$n^{-4} h^{-14} II \xrightarrow{p} \frac{1}{6}\lambda''(x) f^3(x) f'(x) c_2 c_1 (c_2 - c_1^2) \tag{6.5}$$

as $n \to \infty$. The proof of (6.5) is straightforward, hence it is omitted. To prove (6.4), we decompose I as follows:

$$
\begin{aligned}
I &= -\sum_{j=1}^{n} \log \tfrac{1-F_n(X_j)}{1-F(X_j)} k(\tfrac{x-X_j}{h})\{\Delta_{n,1}(x) + (X_j - x)\Delta_{n,2}(x) + (X_j - x)^2\Delta_{n,3}(x)\} \\
&= \sum_{j=1}^{n} \frac{F_n(X_j) - F(X_j)}{1 - F(X_j)}\{1 + O_p(n^{-1/2}\log n)\} \\
&\quad \times k(\tfrac{x-X_j}{h})\{\Delta_{n,1}(x) + (X_j - x)\Delta_{n,2}(x) + (X_j - x)^2\Delta_{n,3}(x)\} \\
&= \tfrac{1+O_p(h)+O_p(n^{-1/2}\log n)}{1-F(x)} \sum_{j=1}^{n}\{F_n(X_j) - F(X_j)\}k(\frac{x-X_j}{h}) \\
&\quad \times\{\Delta_{n,1}(x) + (X_j - x)\Delta_{n,2}(x) + (X_j - x)^2\Delta_{n,3}(x)\} \\
&= \left[\tfrac{1+O_p(h)+O_p(n^{-1/2}\log n)}{1-F(x)}\Delta_{n,1}\sum_{j=1}^{n}\{F_n(X_j) - F(X_j)\}k(\frac{x-X_j}{h})\right] \\
&\quad + \left[\tfrac{1+O_p(h)+O_p(n^{-1/2}\log n)}{1-F(x)}\{\Delta_{n,2} - n^3h^9f^2(x)f'(x)c_1(c_2 - c_1^2)\}\right. \\
&\qquad \left.\times\sum_{j=1}^{n}\{F_n(X_j) - F(X_j)\}(X_j - x)k(\frac{x-X_j}{h})\right] \\
&\quad + \left[\tfrac{1+O_p(h)+O_p(n^{-1/2}\log n)}{1-F(x)}\{\Delta_{n,3}(x) - n^3h^9f(x)[f'(x)]^2c_1(c_1^2 - c_2)\}\right. \\
&\qquad \left.\times\sum_{j=1}^{n}\{F_n(X_j) - F(X_j)\}(X_j - x)^2k(\frac{x-X_j}{h})\right] \\
&\quad + \left[\tfrac{1+O_p(h)+O_p(n^{-1/2}\log n)}{1-F(x)}\sum_{j=1}^{n}\{F_n(X_j) - F(X_j)\}k(\frac{x-X_j}{h})\right. \\
&\qquad \left.\times\{(X_j - x)n^3h^9f^2(x)f'(x)c_1(c_2 - c_1^2) + (X_j - x)^2n^3h^9f(x)[f'(x)]^2c_1(c_1^2 - c_2)\}\right] \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}
\tag{6.6}
$$

Using (6.1)–(6.3), the terms I_1–I_3 in (6.6) satisfy that

$$\sqrt{nh}n^{-4}h^{-12}I_j = o_p(1), \ j = 1, 2, 3. \tag{6.7}$$

Next, the term I_4 in (6.6) can be estimated as follows:

$$
\begin{aligned}
& I_4 \tfrac{1-F(x)}{1+O_p(h)+O_p(n^{-1/2}\log n)} \\
= \;& n\int\{F_n(s)-F(s)\}k(\tfrac{x-s}{h})\{(s-x)n^3h^9f^2(x)f'(x)c_1(c_2-c_1^2) \\
& +(s-x)^2n^3h^9f(x)[f'(x)]^2c_1(c_1^2-c_2)\}\,dF_n(s) \\
= \;& \tfrac{n}{2}\int k(\tfrac{x-s}{h})\{(s-x)n^3h^9f^2(x)f'(x)c_1(c_2-c_1^2) \\
& \quad +(s-x)^2n^3h^9f(x)[f'(x)]^2c_1(c_1^2-c_2)\}\,d[F_n(s)-F(s)]^2 \\
& +n\int\{F_n(s)-F(s)\}k(\tfrac{x-s}{h})\{(s-x)n^3h^9f^2(x)f'(x)c_1(c_2-c_1^2) \\
& \quad +(s-x)^2n^3h^9f(x)[f'(x)]^2c_1(c_1^2-c_2)\}f(s)\,ds \\
= \;& \Bigg[-\tfrac{n}{2}\int[F_n(s)-F(s)]^2\,d\{k(\tfrac{x-s}{h}) \\
& \quad \times[(s-x)n^3h^9f^2(x)f'(x)c_1(c_2-c_1^2)+(s-x)^2n^3h^9f(x)(f'(x))^2c_1(c_1^2-c_2)]\}\Bigg] \\
& +\Bigg[\sqrt{n}\int\{\sqrt{n}[F_n(s)-F(s)]-B_n(F(s))\}k(\tfrac{x-s}{h}) \\
& \quad \times\{(s-x)n^3h^9f^2(x)f'(x)c_1(c_2-c_1^2)+(s-x)^2n^3h^9f(x)[f'(x)]^2c_1(c_1^2-c_2)\}f(s)\,ds\Bigg] \\
& +\Bigg[\sqrt{n}\int B_n(F(s))k(\tfrac{x-s}{h}) \\
& \quad \times\{(s-x)n^3h^9f^2(x)f'(x)c_1(c_2-c_1^2)+(s-x)^2n^3h^9f(x)[f'(x)]^2c_1(c_1^2-c_2)\}f(s)\,ds\Bigg] \\
= \;& III_1+III_2+III_3.
\end{aligned} \tag{6.8}
$$

Since

$$
\begin{aligned}
EIII_3^2 &= 2nh^2 \int_{-1}^1 \int_{-1}^t \{F(x-th) - F(x-th)F(x-sh)\}k(s)k(t) \\
&\quad \times\{-sn^3h^{10}f^2(x)f'(x)c_1(c_2-c_1^2) + s^2n^3h^{11}f(x)[f'(x)]^2c_1(c_1^2-c_2)\} \\
&\quad \times\{-tn^3h^{10}f^2(x)f'(x)c_1(c_2-c_1^2) + t^2n^3h^{11}f(x)[f'(x)]^2c_1(c_1^2-c_2)\} \\
&\quad \times f(x-sh)f(x-th)\,dsdt \\
&= nh^2 \int_{-1}^1 \int_{-1}^1 F(x)[1-F(x)]k(s)k(t) \\
&\quad \times\{-sn^3h^{10}f^2(x)f'(x)c_1(c_2-c_1^2) + s^2n^3h^{11}f(x)[f'(x)]^2c_1(c_1^2-c_2)\} \\
&\quad \times\{-tn^3h^{10}f^2(x)f'(x)c_1(c_2-c_1^2) + t^2n^3h^{11}f(x)[f'(x)]^2c_1(c_1^2-c_2)\} \\
&\quad \times f(x-sh)f(x-th)\,dsdt \\
&\quad +2nh^2 \int_{-1}^1 \int_{-1}^t \{-thf(x) + thF(x)f(x) + shF(x)f(x) + O(h^2)\}k(s)k(t) \\
&\quad \times\{-sn^3h^{10}f^2(x)f'(x)c_1(c_2-c_1^2) + s^2n^3h^{11}f(x)[f'(x)]^2c_1(c_1^2-c_2)\} \\
&\quad \times\{-tn^3h^{10}f^2(x)f'(x)c_1(c_2-c_1^2) + t^2n^3h^{11}f(x)[f'(x)]^2c_1(c_1^2-c_2)\} \\
&\quad \times f(x-sh)f(x-th)\,dsdt \\
&= O(n^7h^{24}) + 2n^7h^{23}f^7(x)[f'(x)]^2c_1^2(c_2-c_1^2)^2 \\
&\quad \times \int_{-1}^1 \int_{-1}^t \{-t + tF(x) + sF(x)\}k(s)k(t)st\,dsdt \\
&= O(n^7h^{24}) + 2n^7h^{23}f^7(x)[f'(x)]^2c_1^2(c_2-c_1^2)^2[F(x)-1] \\
&\quad \times \int_{-1}^1 \int_{-1}^t k(s)k(t)(s+t)st\,dsdt \\
&\quad +2n^7h^{23}f^7(x)[f'(x)]^2c_1^2(c_2-c_1^2)^2 \int_{-1}^1 \int_{-1}^t k(s)k(t)s^2t\,dsdt \\
&= O(n^7h^{24}) + n^7h^{23}f^7(x)[f'(x)]^2c_1^2(c_2-c_1^2)^2[F(x)-1] \\
&\quad \times \int_{-1}^1 \int_{-1}^1 k(s)k(t)(s+t)st\,dsdt \\
&\quad +n^7h^{23}f^7(x)[f'(x)]^2c_1^2(c_2-c_1^2)^2c_3 \\
&= O(n^7h^{24}) + n^7h^{23}f^7(x)[f'(x)]^2c_1^2(c_2-c_1^2)^2c_3,
\end{aligned}
$$

we have

$$\sqrt{nh}n^{-4}h^{-12}III_3 \xrightarrow{d} N\Big(0, f^7(x)[f'(x)]^2c_1^2(c_2 - c_1^2)^2c_3\Big). \tag{6.9}$$

Using (6.1)–(6.3), we notice that terms III_1 and III_2 in (6.8) satisfy

$$\sqrt{nh}n^{-4}h^{-12}III_j = o_p(1),\ j = 1, 2. \tag{6.10}$$

Thus, (6.4) follows from (6.7), (6.9) and (6.10). This completes the proof of Theorem 1.

Proof of Theorem 2. Define

$$H(u) = H_1(u) + H_2(u),$$

with $H_1(u) = P(Z_j \le u, \delta_j = 1)$ and $H_2(u) = P(Z_j \le u, \delta_j = 0)$, and let

$$H_n(u) = H_{n,1}(u) + H_{n,2}(u)$$

with $H_{n,1}(u) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \le u, \delta_j = 1)$ and $H_{n,2}(u) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \le u, \delta_j = 0)$. Let T be such that $1 - H(T) > d$ with some $d > 0$ and M, λ denote generic positive constants. Then it follows from Major and Rejto (1988) that the process $\{F_n^*(u) - F(u), -\infty < u < \infty, 1 - H(u) > 0\}$ can be represented as

$$F_n^*(u) - F(u) = (1 - F(u))[B_1(n, u) + B_2(n, u)] + R(n, u),$$

where

$$B_1(n, u) = \frac{H_{n,1}(u) - H_1(u)}{1 - H(u)} - \int_{-\infty}^{u} \frac{H_{n,1}(y) - H_1(y)}{[1 - H(y)]^2}\, dH(y)\,,$$

$$B_2(n, u) = \int_{-\infty}^{u} \frac{H_n(y) - H(y)}{[1 - H(y)]^2}\, dH_2(y)$$

and for any $\delta_0 > 0$

$$P(A_1) \le Me^{-\lambda h^{-\delta_0}}$$

where $A_1 = \{\sup_{u \le T} n|R(n, u)| > h^{-\delta_0}\}$. Moreover, there exists a Gaussian Process $W(u)$, $-\infty < u < \infty$, with $E(W(u)) = 0$ and covariance

$$E(W(s)W(t)) = \gamma(s) = \int_{-\infty}^{s} [1 - G(u)]^{-1}[1 - F(u)]^{-2}\, dF(u) \tag{6.11}$$

for $-\infty < s \le t < \infty$, such that

$$P(A_2) \le Me^{-\lambda h^{-\delta_0}} \quad \text{and} \quad P(A_3) \le Me^{-\lambda h^{-\delta_0}}, \tag{6.12}$$

where $A_2 = \{\sup_{-\infty<u\le T} \sqrt{n}|\sqrt{n}[B_1(n,u) + B_2(n,u)] - W(u)| > h^{-\delta_0}\}$ and $A_3 = \{\sup_{-\infty<u<\infty} \sqrt{n}|H_n(u) - H(u)| > h^{-\delta_0}\}$. It is easy to check that

$$\left\{\begin{array}{l} n^{-1}h^{-1}s_{n,0}(x) = H_1'(x) + O_p(h^2) \\ n^{-1}h^{-3}s_{n,1}(x) = -H_1''(x)c_1 + O_p(h) \\ n^{-1}h^{-3}s_{n,2}(x) = H_1'(x)c_1 + O_p(h^2) \\ n^{-1}h^{-5}s_{n,3}(x) = -H_1''(x)c_2 + O_p(h) \\ n^{-1}h^{-5}s_{n,4}(x) = H_1'(x)c_2 + O_p(h^2). \end{array}\right. \tag{6.13}$$

By (6.13), we obtain that

$$\left\{\begin{array}{l} \Delta_{n,1}(x) = o_p(n^3h^{11}) \\ \Delta_{n,2}(x) = n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2 - c_1^2) + O_p(n^3h^{10}) \\ \Delta_{n,3}(x) = n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2) + O_p(n^3h^{10}) \\ \Delta_n(x) = n^4h^{12}[H_1'(x)]^3H_1''(x)c_1^2(c_2 - c_1^2) + O_p(n^4h^{13}). \end{array}\right. \tag{6.14}$$

Since

$$\begin{aligned} & \{\hat{\lambda}_n^*(x) - \lambda(x)\}\Delta_n(x) \\ = & \left[\sum_{j=1}^n \{\Lambda_n(Z_j) - \Lambda(Z_j)\}k(\frac{x-Z_j}{h})\{\Delta_{n,1}(x) + (Z_j - x)\Delta_{n,2}(x) + (Z_j - x)^2\Delta_{n,3}(x)\}\right] \\ & + \Bigg[\sum_{j=1}^n \{\Lambda(Z_j) - \Lambda_n(x) - \lambda(x)\lambda(Z_j - x) - \frac{1}{2}\lambda'(x)(Z_j - x)^2\}k(\frac{x-Z_j}{h}) \\ & \times\{\Delta_{n,1}(x) + (Z_j - x)\Delta_{n,2}(x) + (Z_j - x)^2\Delta_{n,3}(x)\}\Bigg] \\ = & I + II, \end{aligned} \tag{6.15}$$

Theorem 2 holds if we can show that

$$\sqrt{nh}n^{-4}h^{-12}I \xrightarrow{d} N\left(0, \frac{f(x)[H_1'(x)]^6[H_1''(x)]^2c_1^2(c_2 - c_1^2)^2c_3}{[1 - F(x)]^2[1 - G(x)]}\right) \tag{6.16}$$

and

$$n^{-4}h^{-14}II \xrightarrow{p} \frac{1}{6}\lambda''(x)[H_1'(x)]^3H_1''(x)c_2c_1(c_2-c_1^2) \tag{6.17}$$

as $n \to \infty$, where $H_1'(x) = [1-G(x)]f(x)$. It is easy to check that (6.17) holds. To prove (6.16), we decompose the term I in (6.15) as follows:

$$\begin{aligned}
I &= -\sum_{j=1}^{n}\log\frac{1-F_n^*(Z_j)}{1-F(Z_j)}k(\frac{x-Z_j}{h})\times\{\Delta_{n,1}(x)+(Z_j-x)\Delta_{n,2}(x)+(Z_j-x)^2\Delta_{n,3}(x)\} \\
&= \sum_{j=1}^{n}\frac{F_n^*(Z_j)-F(Z_j)}{1-F(Z_j)}(1+O_p(n^{-1/2}\log n))k(\frac{x-Z_j}{h}) \\
&\quad \times(\Delta_{n,1}(x)+(Z_j-x)\Delta_{n,2}(x)+(Z_j-x)^2\Delta_{n,3}(x)) \\
&= \tfrac{1+O_p(h)+O_p(n^{-1/2}h^{-\delta_0})}{1-F(x)}\sum_{j=1}^{n}\{F_n^*(Z_j)-F(Z_j)\}k(\frac{x-Z_j}{h}) \\
&\quad \times\{\Delta_{n,1}(x)+(Z_j-x)\Delta_{n,2}(x)+(Z_j-x)^2\Delta_{n,3}(x)\} \\
&= \left[\tfrac{1+O_p(h)+O_p(n^{-1/2}h^{-\delta_0})}{1-F(x)}\Delta_{n,1}\sum_{j=1}^{n}\{F_n^*(Z_j)-F(Z_j)\}k(\frac{x-Z_j}{h})\delta_j\right] \\
&\quad +\left[\tfrac{1+O_p(h)+O_p(n^{-1/2}h^{-\delta_0})}{1-F(x)}\{\Delta_{n,2}(x)-n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)\}\right. \\
&\quad \left.\times\sum_{j=1}^{n}\{F_n^*(Z_j)-F(Z_j)\}(Z_j-x)k(\frac{x-Z_j}{h})\delta_j\right] \\
&\quad +\left[\tfrac{1+O_p(h)+O_p(n^{-1/2}h^{-\delta_0})}{1-F(x)}\{\Delta_{n,3}(x)-n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}\right. \\
&\quad \left.\times\sum_{j=1}^{n}\{F_n^*(Z_j)-F(Z_j)\}(Z_j-x)^2k(\frac{x-Z_j}{h})\delta_j\right] \\
&\quad +\left[\tfrac{1+O_p(h)+O_p(n^{-1/2}h^{-\delta_0})}{1-F(x)}\textstyle\sum_{j=1}^{n}\{F_n^*(Z_j)-F(Z_j)\}k(\frac{x-Z_j}{h})\delta_j\right. \\
&\quad \left.\times\{(Z_j-x)n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+(Z_j-x)^2n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}\right] \\
&= I_1+I_2+I_3+I_4.
\end{aligned} \tag{6.18}$$

Using (6.12)-(6.14) the terms I_1–I_3 can be estimated as follows:

$$\sqrt{nh}n^{-4}h^{-12}I_j = o_p(1),\ \ j = 1, 2, 3. \tag{6.19}$$

Next, we estimate the term I_4 in (6.18). Note that

$$dH_{n,1}(u) = [1 - H(u)]d[B_1(n, u) + B_2(n, u)] + H_1'(u)\,du - \frac{H_n(u) - h(u)}{1 - H(u)}H_2'(u)\,du.$$

We obtain that

$$\begin{aligned}
& I_4\tfrac{1-F(x)}{1+O_p(h)+O_p(n^{-1/2}h^{-\delta_0})} \\
= \ & \sum_{i=1}^{n}\delta_i k(\frac{x - Z_i}{h})R(n, Z_i)\big\{(Z_i - x)n^3h^9[H_1'(x)]^9 H_1''(x)c_1(c_2 - c_1^2) \\
& \quad +(Z_i - x)^2n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2)\big\} \\
& +n\textstyle\int[1 - F(s)][B_1(n, s) + B_2(n, s)]k(\frac{x-s}{h})\big\{(s - x)n^3h^9[H_1'(x)]^2h_1''(x)c_1(c_2 - c_1^2) \\
& \quad +(s - x)^2n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2)\big\}\,dH_{n,1}(s) \\
= \ & \Big[\sum_{i=1}^{n}\delta_i k(\frac{x - Z_i}{h})R(n, Z_i)\{(Z_i - x)n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2 - c_1^2) \\
& \quad +(Z_i - x)^2n^3h^9H_1'9x)[H_1''(x)]^2c_1(c_1^2 - c_2)\}\Big] \\
& +\Big[\tfrac{n}{2}\textstyle\int[1 - F(s)]k(\frac{x-s}{h})\{(s - x)n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2 - c_1^2) + (s - x)^2n^3h^9 \\
& \quad \times H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2)\}[1 - H(s)]\,d[B_1(n, s) + B_2(n, s)]^2\Big] \\
& +\Big[\sqrt{n}\textstyle\int[1 - F(s)]\{\sqrt{n}[B_1(n, s) + B_2(n, s)] - W(s)\}k(\frac{x-s}{h}) \\
& \quad \times\{(s - x)n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2 - c_1^2) \\
& \quad +(s - x)^2n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2)\}H_1'(s)\,ds\Big] \\
& +\Big[\sqrt{n}\textstyle\int[1 - F(s)]W(s)k(\frac{x-s}{h})\{(s - x)n^3h^9[H_1'(x)]^2H_1''(x)c_1(c_2 - c_1^2) \\
& \quad +(s - x)^2n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2)\}H_1'(s)\,ds\Big] \\
& -\Big[n\textstyle\int[1 - F(s)][B_1(n, s) + B_2(n, s)]k(\frac{x-s}{h})\{(s - x)n^3h^9[H_1'(x)]^2H_1''(x) \\
& \quad \times c_1(c_2 - c_1^2) + (s - x)^2n^3h^9H_1'(x)[H_1''(x)]^2c_1(c_1^2 - c_2)\}\tfrac{H_n(u)-H(u)}{1-H(u)}H_2'(u)\,du\Big] \\
= \ & III_1 + III_2 + III_3 + III_4 - III_5.
\end{aligned} \tag{6.20}$$

Using (6.12)-(6.14), the terms III_1, III_2, III_3 and III_5 satisfy

$$\sqrt{nh}n^{-4}h^{-12}III_j = o_p(1),\ \ j = 1, 2, 3, 5. \tag{6.21}$$

Since

$$
\begin{aligned}
& EIII_4^2 \\
= \;& 2nh^2\int_{-1}^1\int_{-1}^t[1-F(x-st)][1-F(x-th)]\gamma(x-th)k(s)k(t) \\
& \times\{-sn^3h^{10}[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+s^2n^3h^{11}H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}H_1'(x-sh) \\
& \times\{-tn^3h^{10}[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+t^2n^3h^{11}H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}H_1'(x-th)\,dsdt \\
= \;& nh^2\int_{-1}^1\int_{-1}^1[1-F(x-sh)][1-F(x-th)]\gamma(x)k(s)k(t) \\
& \quad\times\{-sn^3h^{10}[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+s^2n^3h^{11}H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}H_1'(x-sh) \\
& \quad\times\{-tn^3h^{10}[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+t^2n^3h^{11}H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}H_1'(x-th)\,dsdt \\
& +2nh^2\int_{-1}^1\int_{-1}^t[1-F(x-th)][1-F(x-sh)][-th\gamma'(x)+O(h^2)]k(s)k(t) \\
& \quad\times\{-sn^3h^{10}[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+s^2n^3h^{11}H_1'(x)[H_1''(x)]^2c_1(c_1^2-c_2)\}H_1'(x-sh) \\
& \quad\times\{-tn^3h^{10}[H_1'(x)]^2H_1''(x)c_1(c_2-c_1^2)+t^2n^3h^{11}H_1'(x)[H_1''(x)]^2c_1(c_2-c_1^2)\}H_1'(x-th)\,dsdt \\
= \;& O(n^7h^{24})+2n^7h^{23}\int_{-1}^1\int_{-1}^t[1-F(x)]^2[-t\gamma'(x)]k(s)k(t)st[H_1'(x)]^6[H_1''(x)]^2c_1^2(c_2-c_1^2)^2\,dsdt \\
= \;& O(n^7h^{24})-n^7h^{23}\int_{-1}^1\int_{-1}^1[1-F(x)]^2\gamma'(x)[t+s]k(s)k(t)st[H_1'(x)]^6[H_1''(x)]^2c_1^2(c_2-c_1^2)^2\,dsdt \\
& +2n^7h^{23}\int_{-1}^1\int_{-1}^t[1-f(x)]^2s\gamma'(x)k9s)k(t)st[H_1'(x)]^6[H_1''(x)]^2c_1^2(c_2-c_1^2)^2\,dsdt \\
= \;& O(n^7h^{24})+n^7h^{23}[1-G(x)]^{-1}f(x)[H_1'(x)]^6[H_1''(x)]^2c_1^2(c_2-c_1^2)^2c_3,
\end{aligned}
$$

we have

$$
\sqrt{nh}n^{-4}h^{-12}III_3 \xrightarrow{d} N\Bigg(0,[1-G(x)]^{-1}f(x)[H_1'(x)]^6[H_1''(x)]^2c_1^2(c_2-c_1^2)^2c_3\Bigg). \tag{6.22}
$$

Hence, (6.15) follows from (6.19), (6.21) and (6.22). This completes the proof of Theorem 2.

Proof of Theorem 3. It follows from the proof of Theorem 1 that

$$\begin{aligned}
\sqrt{nh}\{\tilde{\lambda}_n(x)-\lambda(x)\} &= \tfrac{1-\sqrt{2}}{4}\sqrt{nh}\{\hat{\lambda}_n(x-(\sqrt{1/2}+1)\delta h)-\lambda(x-(\sqrt{1/2}+1)\delta h)\} \\
&\quad +\tfrac{1}{2}\sqrt{nh}\{\hat{\lambda}_n(x-\sqrt{1/2}\delta h)-\lambda(x-\sqrt{1/2}\delta h)\} \\
&\quad +\tfrac{1+\sqrt{2}}{4}\sqrt{nh}\{\hat{\lambda}_n(x-(\sqrt{1/2}-1)\delta h)-\lambda(x-(\sqrt{1/2}-1)\delta h)\} \\
&\quad +\tfrac{1-\sqrt{2}}{4}\sqrt{nh}\{\lambda(x-(\sqrt{1/2}+1)\delta h)-\lambda(x)\} \\
&\quad +\tfrac{1}{2}\sqrt{nh}\{\lambda(x-\sqrt{1/2}\delta h)-\lambda(x)\} \\
&\quad +\tfrac{1+\sqrt{2}}{4}\sqrt{nh}\{\lambda(x-(\sqrt{1/2}-1)\delta h)-\lambda(x)\} \\
&= \{\tfrac{1-\sqrt{2}}{4}+\tfrac{1}{2}+\tfrac{1+\sqrt{2}}{4}\}\tfrac{b\lambda''(x)c_2}{6c_1}+\tfrac{c_1(c_2-c_1^2)n^4h^{11}f^3(x)f'(x)}{(1-F(x))\Delta_n(x)} \\
&\quad \times\{\tfrac{1-\sqrt{2}}{4}\sqrt{h}\textstyle\int_{-1}^{1}B_n(F(x-(\sqrt{1/2}+1)\delta h+sh))k(s)s\,ds \\
&\qquad +\tfrac{1}{2}\sqrt{h}\textstyle\int_{-1}^{1}B_n(F(x-\sqrt{1/2}\delta h+sh))k(s)s\,ds \\
&\qquad +\tfrac{1+\sqrt{2}}{4}\sqrt{h}\textstyle\int_{-1}^{1}B_n(F(x-(\sqrt{1/2}-1)\delta h+sh)k(s)s\,ds\}+o_p(1).
\end{aligned}$$

Note that

$$\frac{B_n(F(x+sh))-B_n(F(x))}{\sqrt{h}} \xrightarrow{D} \sqrt{f(x)}W(s)$$

in $D([0,T])$, where $T>0$ and $W(s)$ is a Wiener process, and

$$E\Big\{\big(\int_{-1}^{1}W(s-a)k(s)s\,ds\big)\big(\int_{-1}^{1}W(s-b)k(s)s\,ds\big)\Big\}=c_5(a,b),$$

where $c_5(a,b)$ is defined as in (4.1). We prove Theorem 3 by the above equations.

Proof of Theorem 4. Similar to the proof of Theorem 3.

Acknowledgment. Cheng's research was partially supported by NSC grant NSC-94-2118-M-002-002 and Mathematics Division, National Center for Theoretical Sciences at Taipei. Peng's research was supported in part by NSF grant DMS-0403443. Sun's research was supported in part by NSA grant MDA904-02-10071.

References

[1] K. Chen and S. Lo (1997). On the rate of uniform convergence of the product-limit estimator: strong and weak laws. *Ann. Statist., 25(3), 1050 - 1087.*

[2] M.-Y. Cheng, L. Peng and J.-S. Wu (2005). Reducing variance in univariate smoothing. *Technical report.*

[3] S. Diehl and W. Stute (1988). Kernel density and hazard function estimation in the presence of censoring. *Journal of Multivariate Analysis 25, 299 - 310.*

[4] J. Fan and I. Gijbels (1996). Local Polynomial Modelling and its Applications. *Chapman and Hall, London.*

[5] J. Jiang and K. Doksum (2003). On local polynomial estimation of hazard rates and their derivatives under random censoring. *Mathematical Statistics and Applications: Festschrift for Constance van Eeden, 463 - 481, IMS Lecture Notes Monogr. Ser. 42.*

[6] E.L. Kaplan and P. Meier (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc. 53, 457 - 481.*

[7] J. Komlós, P. Major and G. Tusnády (1975). An approximation of partial sums of independent rv's and sample df, I. *Z. Wahrsch. view. Gebiete 32, 111 - 131.*

[8] R.Y.C. Liu and J. Van Ryzin (1985). A histogram estimator of the hazard rate with censored data. *Ann. Statist. 13(2), 592 - 605.*

[9] S.H. Lo, Y.P. Mack and J.L. Wang (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Prob. Th. rel. Fields 80, 461 - 473.*

[10] P. Major and L. Rejto (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist. 16, 1113 - 1132.*

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

[12] V.K. Murphy (1965). Estimation of jumps, reliability and hazard rate. *Ann. Math. Statist. 36, 1032 - 1040.*

[13] P. Patil (1993). On the least squares cross-validation bandwidth in hazard rate estimation. *Ann. Statist. 21(4), 1792 - 1810.*

[14] P. Patil (1997). Nonparametric hazard rate estimation by orthogonal wavelet methods. *Journal of Statistical Planning and Inference 60, 153 - 168.*

[15] J. Rice and M. Rosenblatt (1976). Estimation of the log survivor function and hazard function. *Sankhya Ser. A 38, 60 - 78.*

[16] H. Schäfer (1985). A note on data-adaptive kernel estimation of the hazard and density function in the random censorship situation. *Ann. Statist. 13(2), 818 - 820.*

[17] N.D. Singpurwalla and M.Y. Wong (1983). Kernel estimators of the failure rate function and density estimation: an analogy. *J. Amer. Statist. Assoc. 78(3), 478 - 481.*

[18] M.A. Tanner (1983). A note on the variable kernel estimator of the hazard function randomly censored data. *Ann. Statist. 11(3), 994 - 998.*

[19] M.A. Tanner and W.H. Wong (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist. 11(3), 989 - 993.*

[20] Q.H. Wang (1999). Some bounds for the error of an estimator of the hazard function with censored data. *Statistics & Probability Letters 44, 319 - 326.*

[21] G.S. Watson and M.R. Leadbetter (1964). Hazard analysis I. *Biometrika 51, 175 - 184.*

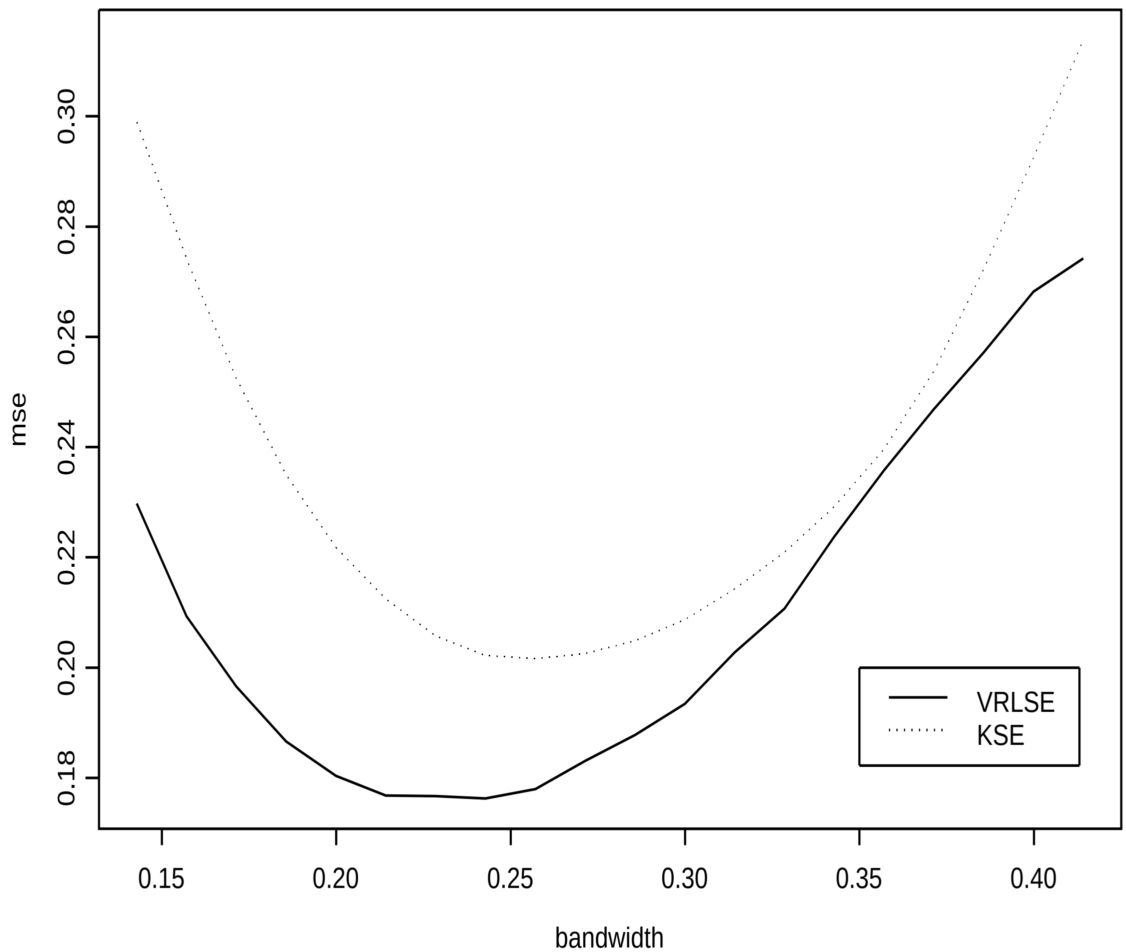

Figure 1: *Mean squared errors.* The mean squared errors of $\bar{\lambda}_n(x)$ (dotted line) and $\tilde{\lambda}_n(x)$ (solid line) are plotted against $h = \frac{\bar{h}_{opt}}{2} + \frac{j}{20}\bar{h}_{opt}, j = 0, 1, \cdots, 19$, where $\bar{h}_{opt}$ is defined in (3.1). We take x such that $F(x) = 0.5$.