Bernoulli 2(2), 1996, 183–198

Some asymptotic results for kernel density estimation under random censorship

BIAO ZHANG

Department of Mathematics, The University of Toledo, Toledo, OH 43606, USA

Random censored data consist of i.i.d. pairs of observations (X_i, δ_i), $i = 1, \ldots, n$. If $\delta_i = 0$, X_i denotes a censored observation, and if $\delta_i = 1$, X_i denotes a survival time, which is the variable of interest. In this paper, we apply the martingale method for counting processes to study asymptotic properties for the kernel estimator of the density function of the survival times. We also derive an asymptotic expression for the mean integrated square error of the kernel density estimator, which can be used to obtain an asymptotically optimal bandwidth.

Keywords: bandwidth; counting process; martingale; Kaplan–Meier estimator; mean integrated square error

1. Introduction

Let $T_1, \ldots, T_n$ be a sequence of independent, non-negative random variables with common continuous distribution function F. Independent of the T_i, let $U_1, \ldots, U_n$ be another sequence of independent, non-negative random variables with common right-continuous distribution function G. We will refer to the T_i as survival times and to the U_i as censoring times. Under the random censorship model, we are only able to observe the smaller of T_i and U_i and an indicator of which variable was smaller:

$$X_i = \min(T_i, U_i), \qquad \delta_i = I_{[T_i \leq U_i]}, \qquad \text{for} \quad i = 1, \ldots n, \tag{1.1}$$

where I_A for any event A, denotes the indicator function of A. Based on these randomly censored data, the Kaplan and Meier (1958) product limit estimator for the survival function F is defined by

$$1 - F_n(t) = \begin{cases} \Pi_{k: X_{(k)} \leq t}\left(1 - \dfrac{\delta_{(k)}}{n-k+1}\right), & \text{if} \quad t \leq X_{(n)}, \\ 1 - F_n(X_{(n)}), & \text{if} \quad t > X_{(n)}, \end{cases} \tag{1.2}$$

where $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ are the order statistics of $X_1, \ldots, X_n$ and $\delta_{(i)}$ is the value of δ associated with $X_{(i)}$, that is, $\delta_{(i)} = \delta_j$ when $X_{(i)} = X_j$. Write $S_n(t) = 1 - F_n(t)$.

In this paper, we assume that the distribution function F has a density function f with respect to Lebesgue measure on $\mathbb{R}$, and we are interested in estimating f using the randomly censored data in (1.1). Based on the Kaplan–Meier estimator F_n, Blum and Susarla (1980)

1350-7265

proposed to estimate f by a sequence of kernel estimators f_n defined by

$$f_n(t) = \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{d}F_n(s), \tag{1.3}$$

where K is a kernel function having finite support on $[-1, 1]$ and h_n is a sequence of positive bandwidths tending to 0 as $n \to \infty$. The properties of the kernel estimator f_n have been examined by Blum and Susarla (1980), Földes *et al.* (1981) and Mielniczuk (1986), among others. It is the purpose of this paper to study the asymptotic properties of f_n using the theory of martingales for counting processes. The martingale approach to the statistical analysis of counting processes was introduced by Aalen (1976; 1977; 1978) and has proved remarkably successful in yielding results about statistical methods for many problems arising in randomly censored data from biomedical studies. Fleming and Harrington (1991) provided an excellent exposition on the counting process and the martingale methods used with censored survival data. In this paper, we apply the counting process approach to establish some asymptotic results of the kernel estimator f_n for arbitrary distribution function G, including the asymptotic bias, weak consistency and asymptotic normality of f_n. Furthermore, we show that this approach can be successfully employed to obtain a simple asymptotic expression for the mean integrated square error of the kernel estimator f_n, from which we can easily derive an asymptotically optimal bandwidth for f_n.

In order to formulate our results, we first introduce some notation. Let $S(t) = 1 - F(t)$, $C(t) = 1 - G(t), \pi(t) = P(X_1 \geq t) = S(t)C(t-), \lambda(t) = f(t)/S(t)$ and $\Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s$. Note that λ and Λ are the hazard function and the cumulative hazard function of the survival times, respectively. Furthermore, the following stochastic processes on $[0, \infty)$ are used throughout this paper:

$$\begin{aligned} N(t) &= \sum_{i=1}^n I_{[X_i \leq t, \delta_i = 1]}, \\ Y(t) &= \sum_{i=1}^n I_{[X_i \geq t]}, \\ A(t) &= \int_0^t Y(u)\mathrm{d}\Lambda(u), \\ M(t) &= N(t) - A(t) = N(t) - \int_0^t Y(u)\lambda(u)\mathrm{d}u. \end{aligned} \tag{1.4}$$

Moreover, we also need the following:

$$\begin{aligned} \Lambda_n(t) &= \int_0^t \frac{\mathrm{d}N(u)}{Y(u)}, \\ \Lambda_n^*(t) &= \int_0^t I_{[Y(u)>0]}\mathrm{d}\Lambda(u), \\ f_n^*(t) &= \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)S_n(s-)\mathrm{d}\Lambda_n^*(s), \\ \tilde{f}_n(t) &= \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)f(s)\mathrm{d}s, \end{aligned} \tag{1.5}$$

where $\Lambda_n(t)$ is the well-known Nelson cumulative hazard estimator of $\Lambda(t)$ (Nelson, 1969). According to Theorem 1.3.1 of Fleming and Harrington (1991, p. 26), the process M given in (1.4) is an $\mathscr{F}_t$-martingale, where $\mathscr{F}_t = \sigma\{I_{[X_i \le t, \delta_i = 1]}, I_{[X_i \le t, \delta_i = 0]} : 0 \le s \le t, i = 1, \ldots, n\}$. In fact, M is a local square-integrable martingale. In this paper, we assume that the kernel function K is bounded on $[-1, 1]$ and satisfies the following conditions:

$$\int_{-1}^{1} K(t)\mathrm{d}t = 1, \qquad \int_{-1}^{1} tK(t)\mathrm{d}t = 0, \qquad \text{and} \qquad \int_{-1}^{1} t^2 K(t)\mathrm{d}t = k_2 \neq 0. \tag{1.6}$$

This paper is structured as follows. In Section 2, we summarize some results about the bias of f_n by expressing $f_n - f_n^*$ as a stochastic integral with respect to the local square-integrable martingale M. Section 3 deals with the weak consistency and the asymptotic normality of f_n. Finally, in Section 4, we derive an asymptotic expression for the mean integrated square error of f_n, including an asymptotically optimal bandwidth.

2. Bias of f_n

Under the notation used in (1.4), the Kaplan–Meier estimator in (1.2) can be expressed as follows (Fleming and Harrington 1991, p. 97):

$$S_n(t) = 1 - F_n(t) = \prod_{s \le t}\left[1 - \frac{\Delta N(s)}{Y(s)}\right] = \prod_{s \le t}[1 - \Delta\Lambda_n(s)], \tag{2.1}$$

where $\Delta N(s) = N(s) - N(s-)$ and $\Delta\Lambda_n(s) = \Lambda_n(s) - \Lambda_n(s-)$. Equation (2.1) implies that

$$S_n(t) = 1 - \int_0^t S_n(s-)\mathrm{d}\Lambda_n(s),$$

so the kernel estimator f_n in (1.2) can be written as

$$f_n(t) = \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right) S_n(s-)\mathrm{d}\Lambda_n(s). \tag{2.2}$$

Simple algebra shows, from (1.5) and (2.2), that

$$\Lambda_n(t) - \Lambda_n^*(t) = \int_0^t \frac{I_{[Y(u) > 0]}}{Y(u)}\mathrm{d}M(u)$$

and

$$f_n(t) - f_n^*(t) = \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right) S_n(s-)\frac{I_{[Y(s) > 0]}}{Y(s)}\mathrm{d}M(s). \tag{2.3}$$

Thus, $f_n(t) - f_n^*(t)$ is a stochastic integral with respect to the local square-integrable martingale $M(t) = N(t) - \int_0^t \lambda(s)Y(s)\mathrm{d}s$. The following theorem summarizes some properties about the bias of f_n.

Theorem 2.1. *Suppose that f is continuous at t.*

(i) *As $n \to \infty$ and $h_n \to 0$,*

$$\tilde{f}_n(t) - f(t) = \int_{-1}^{1} K(u)[f(t - h_n u) - f(t)]\mathrm{d}u \to 0.$$

(ii)

$$\mathrm{E}f_n(t) = \mathrm{E}f_n^*(t) = \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}[S_n(s-)I_{[Y(s)>0]}]\mathrm{d}\Lambda(s).$$

(iii) *If π is positive in a neighbourhood of t, then we have*

$$\mathrm{E}[f_n(t) - \tilde{f}_n(t)] = \mathrm{E}[f_n^*(t) - \tilde{f}_n(t)] = -\frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}\left[I_{[\tau_Y<s]}\frac{S_n(\tau_Y)S(s)}{S(\tau_Y)}\right]\mathrm{d}\Lambda(s),$$

where $\tau_Y = \inf\{s : Y(s) = 0\}$.

(iv) *Under the same conditions as in part (iii), we have, for large n,*

$$|\mathrm{E}[f_n(t) - \tilde{f}_n(t)]| \le \mathrm{e}^{-n\pi(t+h_n)}\int_{-1}^{1}|K(u)|\lambda(t-h_nu)\mathrm{d}u.$$

As a result, $\mathrm{E}[f_n(t) - \tilde{f}_n(t)]$ *converges to zero at an exponential rate as $n \to \infty$ and $h_n \to 0$.*

(v) *In addition to the conditions in part (iii), if f is twice continuously differentiable at t, then as $n \to \infty$ and $h_n \to 0$, we have the following expression for the bias of f_n:*

$$\mathrm{E}f_n(t) - f(t) = \frac{1}{2}h_n^2 f''(t)k_2 + o(h_n^2) + o\left(\frac{1}{n}\right),$$

where k_2 is given in (1.6).

Proof. Part (i) is trivial. For part (ii), using (2.3) and Theorem 2.4.5 of Fleming and Harrington (1991, p. 73) and noting that $\langle M, M\rangle(t) = \int_0^t Y(s)\lambda(s)\mathrm{d}s$, we have

$$\begin{aligned}
\mathrm{E}[f_n(t) - f_n^*(t)]^2 &= \frac{1}{h_n^2}\mathrm{E}\left[\int_0^\infty K\left(\frac{t-s}{h_n}\right)S_n(s-)\frac{I_{[Y(s)>0]}}{Y(s)}\mathrm{d}M(s)\right]^2 \\
&= \frac{1}{h_n^2}\mathrm{E}\int_0^\infty K^2\left(\frac{t-s}{h_n}\right)S_n^2(s-)\frac{I_{[Y(s)>0]}}{Y^2(s)}\mathrm{d}\langle M, M\rangle(s) \\
&= \frac{1}{h_n}\int_{-1}^{1} K^2(u)\mathrm{E}\left[S_n^2[(t-h_nu)-]\frac{I_{[Y(t-h_nu)>0]}}{Y(t-h_nu)}\right]\lambda(t-h_nu)\mathrm{d}u \qquad (2.4) \\
&\le \frac{1}{h_n}\int_{-1}^{1} K^2(u)\lambda(t-h_nu)\mathrm{d}u < \infty,
\end{aligned}$$

and hence

$$\begin{aligned}
\mathrm{E}f_n(t) = \mathrm{E}f_n^*(t) &= \mathrm{E}\left[\frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)S_n(s-)\mathrm{d}\Lambda_n^*(s)\right] \\
&= \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}[S_n(s-)I_{[Y(s)>0]}]\mathrm{d}\Lambda(s).
\end{aligned}$$

To prove parts (iii) and (iv), applying part (ii) and Lemma 3.2.1 of Fleming and Harrington (1991, p. 99) gives

$$
\begin{aligned}
\mathrm{E}[f_n(t) - \tilde{f}_n(t)] &= \mathrm{E}[f_n^*(t) - \tilde{f}_n(t)] \\
&= \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}[S_n(s-)]\mathrm{d}\Lambda(s) - \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}[S_n(s-)I_{[Y(s)=0]}]\mathrm{d}\Lambda(s) \\
&\quad - \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)S(s-)\mathrm{d}\Lambda(s) \\
&= \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}\left[I_{[\tau_Y \le s]}\frac{S_n(\tau_Y)[S(\tau_Y) - S(s-)]}{S(\tau_Y)} - S_n(s-)I_{[\tau_Y < s]}\right]\mathrm{d}\Lambda(s) \\
&= -\frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{E}\left[I_{[\tau_Y < s]}\frac{S_n(\tau_Y)S(s)}{S(\tau_Y)}\right]\mathrm{d}\Lambda(s).
\end{aligned}
$$

As a result, for large n,

$$
\begin{aligned}
|\mathrm{E}[f_n(t) - \tilde{f}_n(t)]| &\le \frac{1}{h_n}\int_0^\infty \left|K\left(\frac{t-s}{h_n}\right)\right|\mathrm{E}\left[I_{[\tau_Y < s]}S_n(\tau_Y)\frac{S(s)}{S(\tau_Y)}\right]\mathrm{d}\Lambda(s) \\
&\le \frac{1}{h_n}\int_0^\infty \left|K\left(\frac{t-s}{h_n}\right)\right|\mathrm{E}\,I_{[\tau_Y < s]}\mathrm{d}\Lambda(s) \\
&= \frac{1}{h_n}\int_0^\infty \left|K\left(\frac{t-s}{h_n}\right)\right|\mathrm{E}\,I_{[Y(s)=0]}\lambda(s)\mathrm{d}s \\
&= \int_{-1}^1 |K(u)|[1 - \pi(t - h_n u)]^n \lambda(t - h_n u)\mathrm{d}u \\
&\le [1 - \pi(t + h_n)]^n \int_{-1}^1 |K(u)|\lambda(t - h_n u)\mathrm{d}u \\
&\le \mathrm{e}^{-n\pi(t+h_n)}\int_{-1}^1 |K(u)|\lambda(t - h_n u)\mathrm{d}u.
\end{aligned}
$$

Part (v) can be derived from parts (i) and (iv) and by a straightforward calculation. The proof is complete. □

Remark 2.1. Theorem 2.1 indicates that the kernel estimator f_n is, in general, not unbiased. However, it is asymptotically unbiased.

Remark 2.2. If the Kaplan–Meier estimator in (1.2) is defined in such a way that the function $S_n(t) = I - F_n(t)$ is equal to zero for $t \ge X_{(n)}$, then in this case the quantity appearing in part (iii) of Theorem 2.1 is identical to zero, because it is proportional to $S_n(\tau_Y)$, where $\tau_Y = X_{(n)}$.

3. Weak consistency and asymptotic normality of f_n

In this section, we establish the weak consistency and asymptotic normality for the kernel density estimator f_n. The following theorem concerns the weak consistency of f_n.

Theorem 3.1. *If f is continuous at t and π is positive in a neighbourhood of t, then, as $n \to \infty$, $h_n \to 0$ and $nh_n \to \infty$,*

$$f_n(t) \xrightarrow{P} f(t).$$

Proof. Since f is continuous at t, $\lambda = f/S$ is continuous at t as well, and hence both f and λ are bounded in a neighbourhood of t. In view of part (i) of Theorem 2.1 and the following inequality,

$$|f_n(t) - f(t)| \leq |f_n(t) - f_n^*(t)| + |f_n^*(t) - \bar{f}_n(t)| + |\bar{f}_n(t) - f(t)|,$$

it is enough to show that, as $n \to \infty$,

$$|f_n(t) - f_n^*(t)| \xrightarrow{P} 0, \tag{3.1}$$

$$|f_n^*(t) - \bar{f}_n(t)| \xrightarrow{P} 0. \tag{3.2}$$

By our equation (2.3) and Corollary 3.4.1 of Fleming and Harrington (1991, p. 113), we have for large n and any $\epsilon, \eta > 0$,

$$\begin{aligned}
P(|f_n(t) - f_n^*(t)| > \epsilon) &= P\left(\left|\frac{1}{h_n}\int_0^\infty K\left(\frac{t-u}{h_n}\right) S_n(u-) \frac{I_{[Y(u)>0]}}{Y(u)} \mathrm{d}M(u)\right| > \epsilon\right) \\
&\leq P\left(\sup_{t-h_n \leq s \leq t+h_n}\left|\frac{1}{h_n}\int_{t-h_n}^{s} K\left(\frac{t-u}{h_n}\right) S_n(u-) \frac{I_{[Y(u)>0]}}{Y(u)} \mathrm{d}M(u)\right| > \epsilon\right) \\
&\leq \frac{\eta}{\epsilon^2} + P\left(\frac{1}{h_n^2}\int_{t-h_n}^{t+h_n} K^2\left(\frac{t-u}{h_n}\right) S_n^2(u-) \frac{I_{[Y(u)>0]}}{Y^2(u)} \mathrm{d}\langle M, M\rangle(u) \geq \eta\right) \\
&= \frac{\eta}{\epsilon^2} + P\left(\frac{1}{h_n}\int_{-1}^{1} K^2(v) S_n^2[(t-h_n v)-] \frac{I_{[Y(t-h_n v)>0]}}{Y(t-h_n v)} \lambda(t-h_n v)\mathrm{d}v \geq \eta\right) \\
&\leq \frac{\eta}{\epsilon^2} + P\left(\frac{1}{nh_n}\frac{1}{n^{-1}Y(t+h_n)} \sup_{s\in[t-h_n, t+h_n]} \lambda(s) \int_{-1}^{1} K^2(v)\mathrm{d}v \geq \eta\right).
\end{aligned}$$

Since η and ϵ are arbitrary, (3.1) is established by the Glivenko–Cantelli theorem applied to

$n^{-1}Y(\cdot)$. To prove (3.2), using (1.5) gives

$$\begin{aligned}
|f_n^*(t)-\tilde{f}_n(t)| &= \left|\frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)[S_n(s-)I_{[Y(s)>0]}-S(s)]\mathrm{d}\Lambda(s)\right| \\
&\le \int_{-1}^{1}|K(u)|I_{[Y(t-h_nu)>0]}|S_n((t-h_nu)-)-S((t-h_nu)-).|\lambda(t-h_nu)\mathrm{d}u \\
&\quad + \int_{-1}^{1}|K(u)|I_{[Y(t-h_nu)=0]}f(t-h_nu)\mathrm{d}u \\
&\le \left[\sup_{s\in[t-h_n,t+h_n]}\lambda(s)\sup_{s\in[t-h_n,t+h_n]}|S_n(s)-S(s)|\right. \\
&\quad \left. + I_{[Y(t+h_n)=0]}\sup_{s\in[t-h_n,t+h_n]}f(s)\right]\int_{-1}^{1}|K(u)|\mathrm{d}u. \qquad (3.3)
\end{aligned}$$

The Glivenko–Cantelli theorem and Theorem 3.4.2 of Fleming and Harrington (1991, p. 115) imply that $\sup_{s\in[t-h_n,t+h_n]}|S_n(s)-S(s)|\xrightarrow{P}0$. Furthermore, it is easy to show that $I_{[Y(t+h_n)=0]}\xrightarrow{P}0$. Thus, (3.2) is established, and hence the proof of Theorem 3.1 is complete. □

Next we study the asymptotic normality of f_n.

Theorem 3.2. *Suppose that f is continuous at t and π is continuous and positive at t. Then, as $n\to\infty$, $h_n\to 0$ and $nh_n\to\infty$,*

$$\sqrt{nh_n}(f_n(t)-\tilde{f}_n(t))\xrightarrow{\mathscr{D}}N(0,\sigma^2(t)),$$

where $\tilde{f}_n(t)$ is given by (1.5), and

$$\sigma^2(t)=\frac{f(t)}{C(t-)}\int_{-1}^{1}K^2(u)\mathrm{d}u. \qquad (3.4)$$

Proof. By (2.3) we can write

$$\begin{aligned}
U_n(t) &\equiv \sqrt{nh_n}(f_n(t)-f_n^*(t)) \\
&= \sqrt{nh_n^{-1}}\int_0^\infty K\left(\frac{t-s}{h_n}\right)S_n(s-)\frac{I_{[Y(s)>0]}}{Y(s)}\mathrm{d}M(s) \\
&= \int_0^\infty H_n(s)\mathrm{d}M(s),
\end{aligned}$$

where

$$H_n(s)=\sqrt{nh_n^{-1}}K\left(\frac{t-s}{h_n}\right)S_n(s-)\frac{I_{[Y(s)>0]}}{Y(s)}.$$

By Theorem 2.4.3 of Fleming and Harrington (1991, p. 70), we have

$$\begin{aligned}\langle U_n, U_n\rangle(t) &= \int_0^\infty H_n^2(s)\mathrm{d}\langle M, M\rangle(s)\\ &= \int_0^\infty H_n^2(s)Y(s)\mathrm{d}\Lambda(s)\\ &= \int_{-1}^1 nK^2(u)S_n^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{Y(t-h_nu)}\lambda(t-h_nu)\mathrm{d}u.\end{aligned}$$

Now, let

$$\begin{aligned}A_1(t) &= \int_{-1}^1 K^2(u)S_n^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{\frac{1}{n}Y(t-h_nu)}\lambda(t)\mathrm{d}u,\\ A_2(t) &= \int_{-1}^1 K^2(u)S_n^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{\pi(t-h_nu)}\lambda(t)\mathrm{d}u,\\ A_3(t) &= \int_{-1}^1 K^2(u)S_n^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{\pi(t)}\lambda(t)\mathrm{d}u,\\ A_4(t) &= \int_{-1}^1 K^2(u)S^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{\pi(t)}\lambda(t)\mathrm{d}u,\\ A_5(t) &= \int_{-1}^1 K^2(u)S^2(t)\frac{I_{[Y(t-h_nu)>0]}}{\pi(t)}\lambda(t)\mathrm{d}u.\end{aligned}$$

Then we can show that

$$\begin{aligned}|\langle U_n, U_n\rangle(t) - A_1(t)| &= \left|\int_{-1}^1 K^2(u)S_n^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{\frac{1}{n}Y(t-h_nu)}[\lambda(t-h_nu)-\lambda(t)]\mathrm{d}u\right|\\ &\le \frac{1}{n^{-1}Y(t+h_n)}\sup_{s\in[t-h_n,t+h_n]}|\lambda(s)-\lambda(t)|\int_{-1}^1 K^2(u)\mathrm{d}u;\end{aligned} \tag{3.5}$$

$$\begin{aligned}|A_1(t)-A_2(t)| &= \left|\int_{-1}^1 K^2(u)S_n^2((t-h_nu)-)I_{[Y(t-h_nu)>0]}\lambda(t)\left[\frac{1}{\frac{1}{n}Y(t-h_nu)}-\frac{1}{\pi(t-h_nu)}\right]\mathrm{d}u\right|\\ &\le \frac{\lambda(t)}{\pi(t+h_n)n^{-1}Y(t+h_n)}\sup_{s\in[t-h_n,t+h_n]}\left|\frac{1}{n}Y(s)-\pi(s)\right|\int_{-1}^1 K^2(u)\mathrm{d}u;\end{aligned} \tag{3.6}$$

$$\begin{aligned}|A_2(t)-A_3(t)| &= \left|\int_{-1}^1 K^2(u)S_n^2((t-h_nu)-)I_{[Y(t-h_nu)>0]}\lambda(t)\left[\frac{1}{\pi(t-h_nu)}-\frac{1}{\pi(t)}\right]\mathrm{d}u\right|\\ &\le \frac{\lambda(t)}{\pi(t)\pi(t+h_n)}\sup_{s\in[t-h_n,t+h_n]}|\pi(s)-\pi(t)|\int_{-1}^1 K^2(u)\mathrm{d}u;\end{aligned} \tag{3.7}$$

$$
\begin{aligned}
|A_3(t)-A_4(t)| &= \left|\int_{-1}^{1} K^2(u)[S_n^2((t-h_nu)-)-S^2((t-h_nu)-)]\frac{I_{[Y(t-h_nu)>0]}}{\pi(t)}\lambda(t)\mathrm{d}u\right| \\
&\le 2\frac{\lambda(t)}{\pi(t)}\sup_{s\in[t-h_n,t+h_n]}|S_n(s)-S(t)|\int_{-1}^{1}K^2(u)\mathrm{d}u; \qquad (3.8)
\end{aligned}
$$

$$
\begin{aligned}
|A_4(t)-A_5(t)| &= \left|\int_{-1}^{1} K^2(u)[S^2(t-h_nu)-S^2(t)]\frac{I_{[Y(t-h_nu)>0]}}{\pi(t)}\lambda(t)\mathrm{d}u\right| \\
&\le 2\frac{\lambda(t)}{\pi(t)}\sup_{s\in[t-h_n,t+h_n]}|S(s)-S(t)|\int_{-1}^{1}K^2(u)\mathrm{d}u; \qquad (3.9)
\end{aligned}
$$

$$
\begin{aligned}
|A_5(t)-\sigma^2(t)| &= \frac{\lambda(t)S^2(t)}{\pi(t)}\int_{-1}^{1}K^2(u)[1-I_{[Y(t-h_nu)>0]}]\mathrm{d}u \\
&\le \frac{f(t)}{C(t-)}I_{[Y(t+h_n)=0]}\int_{-1}^{1}K^2(u)\mathrm{d}u. \qquad (3.10)
\end{aligned}
$$

Combining (3.5)–(3.10), the inequality

$$
\begin{aligned}
|\langle U_n,U_n\rangle(t)-\sigma^2(t)| \le\ & |\langle U_n,U_n\rangle(t)-A_1(t)|+|A_1(t)-A_2(t)|+|A_2(t)-A_3(t)| \\
&+|A_3(t)-A_4(t)|+|A_4(t)-A_5(t)|+|A_5(t)-\sigma^2(t)|
\end{aligned}
$$

and the Glivenko–Cantelli theorem with Theorem 3.4.2 of Fleming and Harrington (1991, p. 115), we have shown under the given conditions that

$$
\langle U_n,U_n\rangle(t)\xrightarrow{\mathrm{P}}\sigma^2(t). \qquad (3.11)
$$

Now, we define

$$
U_{n\epsilon}(t)=\int_0^\infty H_n(s)I_{[|H_n(s)|\ge\epsilon]}\mathrm{d}M(s);
$$

then applying Theorem 2.4.3 of Fleming and Harrington (1991, p. 70) again gives

$$
\begin{aligned}
\langle U_{n\epsilon},U_{n\epsilon}\rangle(t) &= \int_0^\infty H_n^2(s)I_{[|H_n(s)|\ge\epsilon]}Y(s)\mathrm{d}\Lambda(s) \\
&= \int_{-1}^{1}K^2(u)S_n^2((t-h_nu)-)\frac{I_{[Y(t-h_nu)>0]}}{\frac{1}{n}Y(t-h_nu)}\lambda(t-h_nu)I_A^n\mathrm{d}u \\
&\le \frac{1}{n^{-1}Y(t+h_n)}\left[\sup_{s\in[t-h_n,t+h_n]}\lambda(s)\right]I_B^n\int_{-1}^{1}K^2(u)\mathrm{d}u,
\end{aligned}
$$

where A and B are the two sets

$$
A=\left\{\left|K(u)S_n((t-h_nu)-)\frac{I_{[Y(t-h_n)>0]}}{\frac{1}{n}Y(t-h_nu)}\right|\ge\epsilon\sqrt{nh_n}\right\},
$$

and

$$B = \left\{ \frac{\sup_{u\in[-1,1]} |K(u)|}{n^{-1} Y(t+h_n)} \geq \epsilon\sqrt{nh_n} \right\}.$$

Since the Glivenko–Cantelli theorem implies that $I_B(n) \xrightarrow{\mathbf{P}} 0$, we have

$$\langle U_{n\epsilon}, U_{n\epsilon}\rangle(t) \xrightarrow{\mathbf{P}} 0. \tag{3.12}$$

Therefore, it follows from our (3.11) (3.12), and from Theorem 5.1.1 of Fleming and Harrington (1991, p. 204), that

$$U_n(t) = \sqrt{nh_n}(f_n(t) - f_n^*(t)) \xrightarrow{\mathscr{D}} N(0, \sigma^2(t)). \tag{3.13}$$

Next, we are to show

$$\sqrt{nh_n}|f_n^*(t) - \tilde{f}_n(t)| \xrightarrow{\mathbf{P}} 0. \tag{3.14}$$

By (3.3), we have

$$\begin{aligned}\sqrt{nh_n}|f_n^*(t) - \tilde{f}_n(t)| \leq{}& \sqrt{nh_n}\left[\sup_{s\in[t-h_n,t+h_n]} \lambda(s) \sup_{s\in[t-h_n,t+h_n]} |S_n(s) - S(s)|\right] \int_{-1}^{1} |K(u)|\mathrm{d}u \\ &+ \sqrt{nh_n} I_{[Y(t+h_n)=0]} \left[\sup_{s\in[t-h_n,t+h_n]} f(s)\right] \int_{-1}^{1} |K(u)|\mathrm{d}u. \end{aligned} \tag{3.15}$$

It is seen that $\mathrm{E}(\sqrt{nh_n} I_{[Y(t+h_n)=0]}) = \sqrt{nh_n}(1 - \pi(t+h_n))^n$ and $(1-\pi(t+h_n))^n$ converges to zero at an exponential rate since π is assumed to be positive in a neighbourhood of t. Consequently, the second term on the right-hand side of (3.15) converges to zero in probability. On the other hand, similar to the proof of Theorem 6.3.1 of Fleming and Harrington (1991, p. 235), we can show that for any $\epsilon > 0$,

$$\sqrt{n}(S_n(\cdot) - S(\cdot)) \Rightarrow S(\cdot)W(v(\cdot)) \quad \text{on} \quad D[t-\epsilon, t+\epsilon],$$

where $v(t) \equiv \int_0^t [\pi(s)]^{-1} \mathrm{d}\Lambda(s)$ and $W(\cdot)$ is a Wiener process. Thus,

$$\sup_{s\in[t-\epsilon,t+\epsilon]} \{\sqrt{n}|S_n(s) - S(s)|\} \xrightarrow{\mathscr{D}} \sup_{s\in[t-\epsilon,t+\epsilon]} \{S(s)|W(v(s))|\}.$$

So we can conclude for fixed t that as $n \to \infty$,

$$\sup_{s\in[t-\epsilon,t+\epsilon]} \{\sqrt{n}|S_n(s) - S(s)|\} = O_p(1).$$

As a result, as $n \to \infty$ and $h_n \to 0$, we have

$$\sqrt{nh_n} \sup_{s\in[t-h_n,t+h_n]} |S_n(s) - S(s)| = O_p(h_n^{1/2}).$$

Thus, the first term on the right-hand side of (3.15) converges to zero in probability as well, and hence (3.14) holds. The proof of Theorem 3.2 is complete by combining (3.13) with (3.14). □

Corollary 3.1. *In addition to the conditions in Theorem 3.2, if f satisfies the condition $|f(t+h) - f(t)| \leq C|h|^\alpha$ for h in a neighbourhood of* 0, *where $\alpha \in [0,1]$ and C depends only* *on t and if $nh_n^{1+2\alpha} \to 0$ as $n \to \infty$, then we get*

$$\sqrt{nh_n}(f_n(t) - f(t)) \xrightarrow{\mathscr{D}} N(0, \sigma^2(t)),$$

where $\sigma^2(t)$ is given by (3.4).

Proof. As $n \to \infty$,

$$\sqrt{nh_n}|\tilde{f}_n(t) - f(t)| = \sqrt{nh_n}\left|\int_{-1}^{1} K(u)[f(t-h_n u) - f(t)]\mathrm{d}u\right| \le C\sqrt{nh_n^{1+2\alpha}}\int_{-1}^{1} u|K(u)|\mathrm{d}u \to 0.$$

The proof is complete. □

Note that the assumption that $nh_n^{1+2\alpha} \to 0$ for $\alpha \in [0,1]$, as $n \to \infty$, requires that the bandwidth tends to zero faster than $n^{-1/(1+2\alpha)}$, which is not satisfied when h_n is chosen to be the asymptotically optimal bandwidth in (4.11) of the next section. The following corollary gives the asymptotic normality of f_n for bandwidths of order $n^{-1/5}$.

Corollary 3.2. *In addition to the conditions in Theorem 3.2, we assume that f is twice continuously differentiable in a neighbourhood of t and that the bandwidth h_n satisfies $h_n = O(n^{-1/5})$ as $n \to \infty$. Then, we have*

$$\sqrt{nh_n}(f_n(t) - f(t) - \tfrac{1}{2}h_n^2 f''(t)k_2) \xrightarrow{\mathscr{D}} N(0, \sigma^2(t))$$

as $n \to \infty$, where k_2 and σ^2(t) are given by (1.6) *and* (3.4), *respectively.*

Proof. Applying a two-term Taylor expansion gives, as $n \to \infty$,

$$\begin{aligned}\sqrt{nh_n}|\tilde{f}_n(t) - f(t) - \tfrac{1}{2}h_n^2 f''(t)k_2| &= \tfrac{1}{2}\sqrt{n}h_n^{5/2}\left|\textstyle\int_{-1}^{1} u^2K(u)f''(\xi_n)\mathrm{d}u - f''(t)k_2\right| \\ &= \tfrac{1}{2}\sqrt{n}h_n^{5/2}\left|\textstyle\int_{-1}^{1} u^2K(u)[f''(\xi_n) - f''(t)]\mathrm{d}u\right| \\ &\le \tfrac{1}{2}\sqrt{n}h_n^{5/2}\textstyle\int_{-1}^{1} u^2|K(u)||f''(\xi_n) - f''(t)|\mathrm{d}u \to 0,\end{aligned}$$

where $\xi_n \in (\min(t, t - h_n u), \max(t, t - h_n u))$. The proof is complete □

4. Mean integrated square error and optimal bandwidth

In this section, we consider the mean integrated square error of the kernel estimator f_n as a measure of the global accuracy of f_n. The mean integrated square error will enable us to choose an optimal bandwidth. Let τ be such that $\pi(\tau) > 0$; then, for any $\epsilon > 0$, the mean integrated square error of f_n on the interval $[0, \tau - \epsilon]$ is defined to be

$$\mathrm{MISE}(f_n) = \mathrm{E}\left(\int_0^{\tau-\epsilon}[f_n(t) - f(t)]^2\mathrm{d}t\right) \tag{4.1}$$

The following theorem summarizes some properties of $\mathrm{MISE}(f_n)$.

Theorem 4.1. *Suppose that f is continuous on $[0,\tau]$, where τ is such that $\pi(\tau)>0$.*

(i) *As $n\to\infty$ and $h_n\to 0$,*

$$\mathrm{MISE}(f_n)=\int_0^{\tau-\epsilon}[\bar{f}_n(t)-f(t)]^2\mathrm{d}t+\int_0^{\tau-\epsilon}\mathrm{E}[f_n(t)-\bar{f}_n(t)]^2\mathrm{d}t+o\left(\frac{1}{n}\right). \tag{4.2}$$

(ii) *If f is twice continuously differentiable on $[0,\tau]$, then as $n\to\infty$ and $h_n\to 0$, the first term of* (4.2) *can be written as*

$$\int_0^{\tau-\epsilon}[\bar{f}_n(t)-f(t)]^2\mathrm{d}t=\frac{k_2^2}{4}h_n^4\int_0^{\tau-\epsilon}[f''(t)]^2\mathrm{d}t+o(h_n^4),$$

where k_2 is defined in (1.6).

(iii) *If π is continuous on $[0,\tau]$, then as $n\to\infty, h_n\to 0$ and $nh_n\to\infty$, the second term of* (4.2) *can be expressed as*

$$\int_0^{\tau-\epsilon}\mathrm{E}[f_n(t)-\bar{f}_n(t)]^2\mathrm{d}t=\frac{1}{nh_n}\int_{-1}^{1}K^2(t)\mathrm{d}t\int_0^{\tau-\epsilon}\frac{f(t)}{C(t-)}\mathrm{d}t+o\left(\frac{1}{nh_n}\right).$$

(iv) *If f is twice continuously differentiable on $[0,\tau]$ and π is continuous on $[0,\tau]$, then as $n\to\infty$, $h_n\to 0$ and $nh_n\to\infty$, we have*

$$\mathrm{MISE}(f_n)=\frac{k_2^2}{4}h_n^2\int_0^{\tau-\epsilon}[f''(t)]^2\mathrm{d}t+\frac{1}{nh_n}\int_{-1}^{1}K^2(t)\mathrm{d}t\int_0^{\tau-\epsilon}\frac{f(t)}{C(t-)}\mathrm{d}t+o(h_n^4)+o\left(\frac{1}{nh_n}\right). \tag{4.3}$$

Proof. For part (i), since the integrand in (4.1) is non-negative, we have

$$\mathrm{MISE}(f_n)=\int_0^{\tau-\epsilon}[\bar{f}_n(t)-f(t)]^2\mathrm{d}t+\int_0^{\tau-\epsilon}\mathrm{E}[f_n(t)-\bar{f}_n(t)]^2\mathrm{d}t$$
$$+2\int_0^{\tau-\epsilon}[\bar{f}_n(t)-f(t)]\mathrm{E}[f_n(t)-\bar{f}_n(t)]\mathrm{d}t.$$

Using parts (i) and (iv) of Theorem 2.1 gives

$$\left|\int_0^{\tau-\epsilon}\mathrm{E}[\bar{f}_n(t)-f(t)][f_n(t)-\bar{f}_n(t)]\mathrm{d}t\right|$$
$$\le\int_0^{\tau-\epsilon}\left[\int_{-1}^{1}|K(u)||f(t-h_nu)-f(t)|\mathrm{d}u\right]\left[\mathrm{e}^{-n\pi(t+h_n)}\int_{-1}^{1}|K(u)|\lambda(t-h_nu)\mathrm{d}u\right]\mathrm{d}t$$
$$\le\mathrm{e}^{-n\pi(\tau-\epsilon+h_n)}\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}|f(t-h_nu)-f(t)|\right]\left[\int_{-1}^{1}|K(u)|\mathrm{d}u\right]^2\Lambda(\tau-\epsilon+h_n)=o\left(\frac{1}{n}\right).$$

The proof of part (ii) is straightforward and is therefore omitted here. To prove part (iii), we first write

$$\mathrm{E}[f_n(t)-\bar{f}_n(t)]^2=\mathrm{E}[f_n(t)-f_n^*(t)]^2+\mathrm{E}[f_n^*(t)-\bar{f}_n(t)]^2+2\mathrm{E}\{[f_n(t)-f_n^*(t)][f_n^*(t)-\bar{f}_n(t)]\}. \tag{4.4}$$

Now, by our equations (2.4) and Lemma 3.2.1 of Fleming and Harrington (1991, p.99), we have for large n,

$$
\begin{aligned}
&\sup_{t\in[0,\tau-\epsilon]}\left|nh_n\mathrm{E}[f_n(t)-f_n^*(t)]^2-\frac{f(t)}{C(t-)}\int_{-1}^{1}K^2(u)\mathrm{d}u\right| \\
&=\sup_{t\in[0,\tau-\epsilon]}\left|n\int_{-1}^{1}K^2(u)\mathrm{E}\left[S_n^2[(t-h_nu)-]\frac{I_{[Y(t-h_nu)>0]}}{Y(t-h_nu)}\right]\lambda(t-h_nu)\mathrm{d}u-\frac{\lambda(t)S^2(t)}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u\right| \\
&=\sup_{t\in[0,\tau-\epsilon]}\left|\int_{-1}^{1}K^2(u)\mathrm{E}\left[S_n^2[(t-h_nu)-]\left\{\frac{I_{[Y(t-h_nu)>0]}}{n^{-1}Y(t-h_nu)}-\frac{1}{\pi(t-h_nu)}\right\}\right]\lambda(t-h_nu)\mathrm{d}u\right. \\
&\quad+\int_{-1}^{1}K^2(u)\mathrm{E}(S_n^2[(t-h_nu)-])\left[\frac{1}{\pi(t-h_nu)}-\frac{1}{\pi(t)}\right]\lambda(t-h_nu)\mathrm{d}u \\
&\quad+\int_{-1}^{1}K^2(u)\frac{\mathrm{E}(S_n^2[(t-h_nu)-])}{\pi(t)}[\lambda(t-h_nu)-\lambda(t)]\mathrm{d}u \\
&\quad+\int_{-1}^{1}K^2(u)\frac{\mathrm{E}\{S_n^2[(t-h_nu)-]-S^2[(t-h_nu)-]\}}{\pi(t)}\lambda(t)\mathrm{d}u \\
&\quad\left.+\int_{-1}^{1}K^2(u)\frac{S^2(t-h_nu)-S^2(t)}{\pi(t)}\lambda(t)\mathrm{d}u\right| \\
&\le\left[\sup_{s\in[0,\tau]}\mathrm{E}\left|\frac{I_{[Y(s)>0]}}{n^{-1}Y(s)}-\frac{1}{\pi(s)}\right|\right]\left[\sup_{s\in[0,\tau]}\lambda(s)\right]\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}\left|\frac{1}{\pi(t-h_nu)}-\frac{1}{\pi(t)}\right|\right]\left[\sup_{s\in[0,\tau]}\lambda(s)\right]\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}|\lambda(t-h_nu)-\lambda(t)|\right]\frac{1}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+2\left[\sup_{s\in[0,\tau]}\mathrm{E}\{S_n(s-)-S(s-)\}\right]\frac{\lambda(t)}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+2\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}|S(t-h_nu)-S(t)|\right]\frac{\lambda(t)}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\le\left[\sup_{s\in[0,\tau]}\mathrm{E}\left|\frac{I_{[Y(s)>0]}}{n^{-1}Y(s)}-\frac{I_{[Y(s)>0]}}{\pi(s)}\right|+\frac{[1-\pi(\tau)]^n}{\pi(\tau)}\right]\left[\sup_{s\in[0,\tau]}\lambda(s)\right]\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}\left|\frac{1}{\pi(t-h_nu)}-\frac{1}{\pi(t)}\right|\right]\left[\sup_{s\in[0,\tau]}\lambda(s)\right]\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}|\lambda(t-h_nu)-\lambda(t)|\right]\frac{1}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u+2[1-\pi(\tau)]^n\frac{\lambda(t)}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u \\
&\quad+2\left[\sup_{t\in[0,\tau-\epsilon]}\sup_{u\in[-1,1]}|S(t-h_nu)-S(t)|\right]\frac{\lambda(t)}{\pi(t)}\int_{-1}^{1}K^2(u)\mathrm{d}u. \qquad (4.5)
\end{aligned}
$$

Since $Y(s)$ has a binomial distribution with parameters n and $\pi(s)$ for each fixed $s \in [0, \tau]$ and since

$$\begin{aligned}\mathrm{E}\left|\frac{I_{[Y(s)>0]}}{n^{-1}Y(s)} - \frac{I_{[Y(s)>0]}}{\pi(s)}\right| &= \mathrm{E}I_{[Y(s)>0]}\mathrm{E}\left[\left|\frac{1}{n^{-1}Y(s)} - \frac{1}{\pi(s)}\right| \middle| Y(s) > 0\right] \\ &= [1 - [1-\pi(s)]^n]\mathrm{E}\left[\left|\frac{1}{n^{-1}Y(s)} - \frac{1}{\pi(s)}\right| \middle| Y(s) > 0\right],\end{aligned}$$

it follows from Lemma 4.2 of Aalen (1976, p. 18) that

$$\lim_{n\to\infty} \sup_{s\in[0,\tau]} \mathrm{E}\left|\frac{I_{[Y(s)>0]}}{n^{-1}Y(s)} - \frac{I_{[Y(s)>0]}}{\pi(s)}\right| = 0. \tag{4.6}$$

By (4.5), (4.6) and the continuity of f and π, we have shown that

$$\lim_{n\to\infty} \sup_{t\in[0,\tau-\epsilon]}\left| nh_n\mathrm{E}[f_n(t) - f_n^*(t)]^2 - \frac{f(t)}{C(t-)}\int_{-1}^{1} K^2(u)\mathrm{d}u\right| = 0.$$

Thus, we have

$$E[f_n(t) - f_n^*(t)]^2 = \frac{1}{nh_n}\frac{f(t)}{C(t-)}\int_{-1}^{1} K^2(u)\mathrm{d}u + o\left(\frac{1}{nh_n}\right), \tag{4.7}$$

uniformly for $t \in [0, \tau - \epsilon]$. To deal with the second term on the right-hand side of (4.4), let $B(s) = E[I_{[\tau_Y < s]}S_n(\tau_Y)\{S(\tau_Y) - S(s)\}/S(\tau_Y)]$, where $\tau_Y = \inf\{s : Y(s) = 0\}$. Then by Equation (2.13) of Fleming and Harrington (119, p. 104), we can show that for $s \in [0, \tau]$,

$$\begin{aligned}&n\mathrm{E}[S_n(s) - S(s) - B(s)]^2 \\ &= nS^2(s)\int_0^s \mathrm{E}\left[\frac{S_n^2(w-)}{S^2(w)}\frac{I_{[Y(w)>0]}}{Y(w)}\right]\lambda(w)\mathrm{d}w \\ &= S^2(s)\int_0^s \mathrm{E}\left[S_n^2(w-)\left\{\frac{I_{[Y(w)>0]}}{n^{-1}Y(w)} - \frac{I_{[Y(w)>0]}}{\pi(w)}\right\}\right]\frac{\lambda(w)}{S^2(w)}\mathrm{d}w \\ &\quad + S^2(s)\int_0^s \mathrm{E}\left[\frac{S_n^2(w-)I_{[Y(w)>0]}}{\pi(w)}\right]\frac{\lambda(w)}{S^2(w)}\mathrm{d}w \\ &\le \left\{\sup_{w\in[0,\tau]}\mathrm{E}\left[\left|\frac{I_{[Y(w)>0]}}{n^{-1}Y(w)} - \frac{I_{[Y(w)>0]}}{\pi(w)}\right|\right]\right\}\frac{\tau}{S^2(\tau)}\left[\sup_{w\in[0,\tau]}\lambda(w)\right] + \frac{\tau}{\pi(\tau)S^2(\tau)}\left[\sup_{w\in[0,\tau]}\lambda(w)\right],\end{aligned}$$

which, together with (4.6) and the boundness of λ on $[0, \tau]$, implies that

$$\sup_{s\in[0,\tau]} \mathrm{E}[S_n(s) - S(s) - B(s)]^2 = O\left(\frac{1}{n}\right). \tag{4.8}$$

Furthermore, by our equations (1.5) and Lemma 3.2.1 of Fleming and Harrington (1991, p. 99), we can show that for large n,

$$\begin{aligned}
&\sup_{t\in[0,\tau-\epsilon]} \mathrm{E}[f_n^*(t)-\tilde{f}_n(t)]^2 \\
&\quad= \sup_{t\in[0,\tau-\epsilon]}\left\{\mathrm{E}\left[\int_{-1}^{1} K(u)I_{[Y(t-h_nu)>0]}[S_n((t-h_nu)-)-S((t-h_nu)-)]\lambda(t-h_nu)\mathrm{d}u\right.\right.\\
&\qquad\left.\left.-\int_{-1}^{1} K(u)I_{[Y(t-h_nu)=0]}f(t-h_nu)\mathrm{d}u\right]^2\right\}\\
&\le \sup_{t\in[0,\tau-\epsilon]}\left\{4\int_{-1}^{1} K^2(u)\mathrm{E}[S_n((t-h_nu)-)-S((t-h_nu)-)]^2\lambda^2(t-h_nu)\mathrm{d}u\right.\\
&\quad\left.+4\int_{-1}^{1} K^2(u)\mathrm{E}I_{[Y(t-h_nu)=0]}f^2(t-h_nu)\mathrm{d}u\right\}\\
&\le 16\int_{-1}^{1} K^2(u)\mathrm{E}[S_n((t-h_nu)-)-S((t-h_nu)-)-B((t-h_nu)-)]^2\lambda^2(t-h_nu)\mathrm{d}u\\
&\quad+16\int_{-1}^{1} K^2(u)\mathrm{E}[B^2((t-h_nu)-)]\lambda^2(t-h_nu)\mathrm{d}u+4\int_{-1}^{1} K^2(u)\mathrm{E}I_{[Y(t-h_nu)=0]}f^2(t-h_nu)\mathrm{d}u\\
&\le \left[16\sup_{s\in[0,\tau]}\lambda^2(s)\right]\left\{\sup_{s\in[0,\tau]}\mathrm{E}[S_n(s)-S(s)-B(s)]^2\right\}\int_{-1}^{1}K^2(u)\mathrm{d}u\\
&\quad+\left[\sup_{s\in[0,\tau]}P(Y(s)=0)\right]\left[16\sup_{s\in[0,\tau]}\lambda^2(s)+4\sup_{s\in[0,\tau]}f^2(s)\right]\int_{-1}^{1}K^2(u)\mathrm{d}u\\
&\le \left[16\sup_{s\in[0,\tau]}\lambda^2(s)\right]\left\{\sup_{s\in[0,\tau]}\mathrm{E}[S_n(s)-S(s)-B(s)]^2\right\}\int_{-1}^{1}K^2(u)\mathrm{d}u\\
&\quad+\mathrm{e}^{-n\pi(\tau)}\left[16\sup_{s\in[0,\tau]}\lambda^2(s)+4\sup_{s\in[0,\tau]}f^2(s)\right]\int_{-1}^{1}K^2(u)\mathrm{d}u.
\end{aligned}$$

As a result, by (4.8) and the boundness of f and λ on $[0,\tau]$,

$$\sup_{t\in[0,\tau-\epsilon]} \mathrm{E}[f_n^*(t)-\tilde{f}_n(t)]^2 = O\left(\frac{1}{n}\right). \tag{4.9}$$

Finally, by (4.7), (4.9), and the Schwarz inequality, we have

$$\sup_{t\in[0,\tau-\epsilon]} \mathrm{E}\{[f_n(t)-f_n^*(t)][f_n^*(t)-\tilde{f}_n(t)]\} = O\left(\frac{1}{n\sqrt{h_n}}\right). \tag{4.10}$$

Therefore, combining (4.4), (4.7), (4.9) and (4.10) completes the proof of part (iii). Part (iv) is easily derived from parts (i), (ii) and (iii). The proof of Theorem 4.1 is complete. □

In view of Theorem 4.1, the ideal choice of the bandwidth h_n, from the point of view of minimizing the mean integrated square error of f_n, can be derived by minimizing the sum of the two leading terms in (4.3). In doing so, we have arrived at the following results regarding the asymptotically optimal bandwidth and the corresponding mean integrated square error of f_n.

Corollary 4.1. *The asymptotically optimal bandwidth is*

$$h_n^* = \frac{1}{n^{1/5}} \frac{1}{k_2^{2/5}} \left\{ \int_{-1}^{1} K^2(t)\mathrm{d}t \int_0^{\tau-\epsilon} \frac{f(t)}{C(t-)}\mathrm{d}t \right\}^{1/5} \left\{ \int_0^{\tau-\epsilon} [f''(t)]^2 \mathrm{d}t \right\}^{-1/5}. \tag{4.11}$$

Substituting h_n^* into (4.3), *we have the following expression for the mean integrated square error of f_n with optimal bandwidth h_n^*:*

$$\mathrm{MISE}(f_n) = \frac{5}{4} \frac{k_2^{2/5}}{n^{4/5}} \left\{ \int_{-1}^{1} K^2(t)\mathrm{d}t \int_0^{\tau-\epsilon} \frac{f(t)}{C(t-)}\mathrm{d}t \right\}^{4/5} \left\{ \int_0^{\tau-\epsilon} [f''(t)]^2 \mathrm{d}t \right\}^{1/5} + o(n^{-4/5}). \quad \square$$

Acknowledgement

The author wishes to thank two referees and a case editor for a careful reading of and many valuable comments and constructive suggestions on an earlier version of this paper, which improved the results and presentation.

References

Aalen, O.O. (1976) Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.*, **3**, 15–27.

Aalen, O.O. (1977) Weak convergence of stochastic integrals related to counting processes. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **38**, 261–277.

Aalen, O.O. (1978) Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701–726.

Blum, J.R. and Susarla, V. (1980) Maximal deviation theory of density and failure rate function estimates based on censored data. In P.K. Krishnaiah (ed.), *Multivariate Analysis*, Vol. 5, pp. 213–222. New York: North-Holland.

Fleming, T.R. and Harrington, D.P. (1991) *Counting Processes and Survival Analysis*. New York: John Wiley & Sons.

Földes, A., Rejtö, L. and Winter, B.B. (1981) Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate. *Period. Math. Hungar.*, **12**, 15–29.

Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481, 562–563.

Mielniczuk, J. (1986) Some asymptotic properties of kernel estimators of a density function in case of censored data. *Ann. Statist.*, **14**, 766–773.

Nelson, W. (1969) Hazard plotting for incomplete failure data. *J. Qual. Technol.*, **1**, 27–52.

Received September 1994 and revised October 1995