Annals of the Institute of Statistical Mathematics (2020) 72:187–211
https://doi.org/10.1007/s10463-018-0685-6

New kernel estimators of the hazard ratio and their asymptotic properties

Taku Moriyama[1] · **Yoshihiko Maesono**[2]

Received: 13 September 2017 / Revised: 18 July 2018 / Published online: 21 August 2018

Abstract
We propose a kernel estimator of a hazard ratio that is based on a modification of Ćwik and Mielniczuk (Commun Stat-Theory Methods 18(8):3057–3069, 1989)'s method. A naive nonparametric estimator is Watson and Leadbetter (Sankhyā: Indian J Stat Ser A 26(1):101–116, 1964)'s one, which is naturally given by the kernel density estimator and the empirical distribution estimator. We compare the asymptotic mean squared error ($AMSE$) of the hazard estimators, and then, it is shown that the asymptotic variance of the new estimator is usually smaller than that of the naive one. We also discuss bias reduction of the proposed estimator and derived some modified estimators. While the modified estimators do not lose nonnegativity, their $AMSE$ is small both theoretically and numerically.

Keywords Kernel estimator · Hazard ratio · Nonparametric estimator · Mean squared error

1 Introduction

Rosenblatt (1956) proposed a kernel smoothed estimator of the probability density function $f(\cdot)$. Many researchers have since developed various nonparametric estimators for probability distribution, regression function, hazard ratio, etc. Most of the kernel estimators are inferior in convergence rate, and many researchers have attempted

✉ Taku Moriyama
t-moriyama@imi.kyushu-u.ac.jp

Yoshihiko Maesono
maesono@math.kyushu-u.ac.jp

1 Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

2 Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

to reduce the asymptotic mean squared error ($AMSE$). Although there are many bias reduction methods (e.g., Ruppert and Cline 1994's transformation, Jones et al. 1995's extrapolation, Terrell and Scott 1980's extrapolation for density estimation, which we will discuss in Sect. 4), variance reduction is quite difficult. In density ratio estimation, Ćwik and Mielniczuk (1989) proposed a kernel estimator that they called 'direct.' The $AMSE$ of the direct estimator is different from the $AMSE$ of the naive nonparametric estimator. In this paper, we devise a 'direct' estimator of the hazard ratio by modifying Ćwik and Mielniczuk (1989)'s method, and discuss its $AMSE$.

First, we will describe the direct estimator of the density ratio. Let $X_1, X_2, \ldots, X_n$ be independently and identically distributed ($i.i.d.$) random variables with a distribution function $F(\cdot)$, and $Y_1, Y_2, \ldots, Y_n$ be $i.i.d.$ random variables with a distribution function $G(\cdot)$. $f(\cdot)$ and $g(\cdot)$ are the density functions of $F(\cdot)$ and $G(\cdot)$, and we assume that $g(x_0) \neq 0$ ($x_0 \in \mathbb{R}$). A naive estimator of the density ratio $f(x_0)/g(x_0)$ at the point x_0 is given by $\widehat{f}(x_0)/\widehat{g}(x_0)$ where

$$\widehat{f}(x_0) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x_0 - w}{h}\right) \mathrm{d}F_n(w)$$

and

$$\widehat{g}(x_0) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x_0 - z}{h}\right) \mathrm{d}G_n(z).$$

$K(\cdot)$ is a kernel function, h is a bandwidth that satisfies $h \to 0$ and $nh \to \infty$ ($n \to \infty$), and $F_n(\cdot)$ and $G_n(\cdot)$ are the empirical distribution functions of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$, respectively. We call $\widehat{f}(x_0)/\widehat{g}(x_0)$ an 'indirect' estimator. Ćwik and Mielniczuk (1989) proposed a direct estimator, given by

$$\widehat{\frac{f}{g}}(x_0) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{G_n(x_0) - G_n(w)}{h}\right) \mathrm{d}F_n(w).$$

Chen et al. (2009) obtained an explicit form of its $AMSE$.

In this paper, we develop a new 'direct' estimator of the hazard ratio function by modifying Ćwik and Mielniczuk (1989)'s method and investigate its $AMSE$ (in Sect. 2). We compare the naive and direct estimators (in Sect. 3) and find that our direct estimator performs asymptotically better especially in exponential or gamma cases, which play a central role in survival analysis. Although the bias of the direct estimator is large in some cases, the asymptotic variance is always small when both bandwidth parameters are the same. As mentioned before, there are many bias reduction methods; we discuss them in Sect. 4. We derived some modified estimators, and it is shown that the modified estimators performs well both theoretically and numerically. Proofs of the theorems herein are given in the 'Appendices'.

2 Hazard ratio estimators and their asymptotic properties

2.1 Hazard ratio estimators

Let us assume that the density function $f(\cdot)$ of X_i satisfies $f(x_0) \neq 0$ ($x_0 \in \mathbf{R}$). The hazard ratio function is a type of relative risk and is defined as

$$H(x_0) = \frac{f(x_0)}{1 - F(x_0)}.$$

The meaning of $H(x)\mathrm{d}x$ is the conditional probability of 'death' in $[x, x + \mathrm{d}x]$ given survival to x, and this is a fundamental measure of the difference between several risk groups. The hazard ratio also uniquely determines the 'survival function,' as follows:

$$S(x) = exp\left(\int_{-\infty}^{x} H(u)\mathrm{d}u\right),$$

which gives the probability that a person survives longer than x. These estimators have been extensively discussed over the years, and the Kaplan–Meier and Nelson–Aalen estimators are widely known. Though they are discrete, we can construct a smoothed hazard estimator by using the kernel method. If there is no censoring, the smoothed hazard estimator coincides with the naive estimator, which we will define later.

The estimator of H is useful for describing and testing the effects of medicine, covariates, and so on. Actuaries call it the 'force of mortality' and use it to estimate insurance payouts. In reliability theory, it is called the 'intensity function' and used to evaluate tolerance. The gamma and Weibull forms are typical models of the intensity function, and they describe various random behaviors. In extreme value theory, the hazard ratio determines the form of the extreme value distribution (see Gumbel 1958), which is defined as

$$G_\gamma(x) = exp\left(-(1 + \gamma x)^{-1/\gamma}\right) \quad (1 + \gamma x > 0)$$

where γ is real and called the extreme value index. Let F be a distribution function and x^* be its right endpoint. Under some regularity conditions, if

$$\lim_{x \uparrow x^*} \left(\frac{1}{H(x)}\right)' = \gamma$$

holds, then F is in the domain of attraction of G_γ [i.e. the distribution of a suitably standardized sample maximum converges to G_γ (see De Haan and Ferreira 2007)].

There are also many parametric models describing the dependency of covariates; the most popular one is Cox's proportional hazard model. For the sake of simplicity, we will not consider covariates and instead focus on nonparametric estimation of the baseline hazard. The naive nonparametric estimator of $H(x_0)$ is given by Watson and Leadbetter (1964)

$$\widetilde{H}(x_0) = \frac{\widehat{f}(x_0)}{1 - F_n(x_0)},$$

where

$$\widehat{f}(x_0) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{x_0 - w}{h}\right) \mathrm{d}F_n(w),$$

$K(\cdot)$ is the kernel function, and $F_n(\cdot)$ is the empirical distribution function. By using the properties of the kernel density estimator, Murthy (1965) proved the consistency and asymptotic normality of $\widetilde{H}(x_0)$. Tanner and Wong (1983) proved these properties in the random censorship model by using Hájek's projection method. Patil (1993) gave its mean integrated squared error ($MISE$) and discussed the cross-validation method for selecting the optimal bandwidth in both uncensored and censored settings. Müller and Wang (1994) discussed reduction of boundary bias by using varying kernels methods for nonnegative data. For dependent data, Quintela-del Río (2007) obtained the MSE of the indirect estimator. By using Vieu (1991b)'s results, he obtained a modified $MISE$ that avoids any chance of the denominator being equal to 0. In this paper, we assume that the support of the kernel $K(\cdot)$ is given by a closed interval and there is no censoring.

By extending the idea of Ćwik and Mielniczuk (1989), we develop a new 'direct' estimator of the hazard ratio function, as follows:

$$\widehat{H}(x_0) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{M_n(x_0) - M_n(w)}{h}\right) \mathrm{d}F_n(w),$$

where

$$M_n(w) = w - \int_{-\infty}^{w} F_n(u)\mathrm{d}u = \frac{1}{n}\sum_{i=1}^{n} w - (w - X_i)_+$$

and $(x)_+ = x$ (for $x \geq 0$), $= 0$ (for $x \leq 0$). Though the proposed estimator is not a ratio in appearance, $\widehat{H}(x_0)$ coincides with the following statistic:

$$\widehat{H}(x_0) = \frac{\widehat{f}_{M_n(X)}(x_0)}{M_n'(x_0)} = \frac{\widehat{f}_{M_n(X)}(x_0)}{1 - F_n(x_0)},$$

where $\widehat{f}_{M_n(X)}$ denotes a kernel density estimator for X, which uses an 'empirical transformation' $(X \mapsto M_n(X))$ (see Ruppert and Cline 1994). Therefore, the difference between $\widehat{H}$ and $\widetilde{H}$ is whether we use a transformation or not in kernel density estimation. Since

$$\widehat{f}_{M_n(X)}(x_0) = \frac{M_n'(x_0)}{h}\int_{-\infty}^{\infty} K\left(\frac{M_n(x_0) - M_n(w)}{h}\right) \mathrm{d}F_n(w),$$

the proposed estimator is no longer ratio. Since the asymptotic variance of the proposed estimator is

$$V[\widehat{H}(x_0)] \approx (nh)^{-1} A_{2,0} H(x_0),$$

the asymptotic variance of $\widehat{H}(x_0)$ is much smaller when $H(x_0)$ is large. Ruppert and Cline (1994) recommended transforming X to an (asymptotically) uniform random variable ($X \mapsto \widehat{F}(X)$) so that the transform-kernel density estimator reduces the asymptotic bias. In fact, the hazard estimator

$$\widehat{H}_{RC}(x_0) = \frac{\widehat{f}_{\widehat{F}(X)}(x_0)}{1 - F_n(x_0)}$$

has an improved convergence rate; however, the finite sample performance is not stable and rather poor in many cases as shown in Sect. 4.

Asymptotic properties of the proposed estimator $\widehat{H}(\cdot)$ is discussed below.

2.2 Asymptotic properties

For the sake of simplicity, we will use the notation,

$$A_{i,j} = \int_{-\infty}^{\infty} K^i(u) u^j \mathrm{d}u.$$

The proofs of the theorems are in the 'Appendices.' In hazard ratio estimation, the following holds for fixed x_0 under some regularity conditions:

$$E\left[\left(\check{H}(x_0) - H(x_0)\right)^2\right] \approx \left(\frac{1}{1-F}(x_0) E\left[\check{f}(x_0) - f(x_0)\right]\right)^2 + \frac{1}{(1-F)^2}(x_0) V[\check{f}(x_0)], \quad (1)$$

where

$$\check{H}(x_0) = \frac{\check{f}(x_0)}{1 - F_n(x_0)}.$$

For the transformed density estimator $\widehat{f}_{M(X)}(x_0)$, we have the following $AMSE$:

$$E\left[\left(\widehat{f}_{M(X)}(x_0) - f(x_0)\right)^2\right] = \frac{h^4}{4} \frac{A_{1,2}^2}{m^6(x_0)} \left[f''(x_0) - \frac{f m''}{m}(x_0) - 3\frac{f' m'}{m}(x_0) + 3\frac{f(m')^2}{m^2}(x_0)\right]^2 + \frac{A_{2,0}}{nh} f(x_0) m(x_0),$$

where $M(x) = x - \int_{-\infty}^{x} F(u)\mathrm{d}u$ and $m(x) = M'(x) = 1 - F(x)$. For the direct hazard estimator $\widehat{H}$, we have the following $AMSE$.

Theorem 1 *Let us assume that* (i) $f(\cdot)$ *is three-times differentiable at* x_0 *and* $f^{(3)}(x_0)$ *is bounded,* (ii) K *is symmetric and the support is given by a closed interval,* (iii) $K^{(3)}$ *is bounded, and* (iv) $A_{1,4}$ *and* $A_{2,0}$ *are bounded. Then, the* MSE *of* $\widehat{H}(x_0)$ *is given by*

$$\begin{aligned} &E\left[\left(\widehat{H}(x_0) - H(x_0)\right)^2\right] \\ &= \frac{h^4}{4}A_{1,2}^2\left[\frac{\{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3\}^2}{(1-F)^{10}}\right](x_0) + \frac{A_{2,0}}{nh}H(x_0) \\ &\quad + O\left(h^6 + \frac{1}{nh^{1/2}}\right). \end{aligned} \tag{2}$$

Remark 1 In order to get the above approximations, we perform a Taylor expansion of the integral. We can divide the integral at discrete points, so we do not need to worry about the differentiability of the density function at finite points.

On the other hand, under some regularity conditions, Patil (1993) gave the MSE of $\widetilde{H}(x_0)$, as follows:

$$\begin{aligned} &E\left[\left(\widetilde{H}(x_0) - H(x_0)\right)^2\right] \\ &= \frac{h^4}{4}A_{1,2}^2\left[\frac{(f'')^2}{(1-F)^2}\right](x_0) + \frac{A_{2,0}}{nh}\left[\frac{H}{1-F}\right](x_0) \\ &\quad + O\left(h^6 + \frac{1}{nh^{1/2}}\right). \end{aligned} \tag{3}$$

The asymptotic variances are the second terms on the right hand side of (2) and (3), and the direct estimator has a small variance because of $0 < 1 - F(x_0) < 1$ when both bandwidth parameters are the same. By minimizing the leading terms in the $AMSE$, we have an optimal bandwidth $h = h^*$ of $\widehat{H}(x_0)$, where

$$h^* = n^{-1/5}\left(\frac{A_{2,0}}{A_{1,2}^2}\left[\frac{(1-F)^9 f}{\{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3\}^2}\right](x_0)\right)^{1/5}.$$

In the same way, the optimal bandwidth $h = h^{**}$ of $\widetilde{H}(x_0)$ is given by

$$h^{**} = n^{-1/5}\left(\frac{A_{2,0}}{A_{1,2}^2}\frac{f}{(f'')^2}(x_0)\right)^{1/5}.$$

Furthermore, we can show the asymptotic normality of the direct $\widehat{H}$.

Theorem 2 *Suppose that Theorem* 2.1 *holds. When* $h = cn^{-\xi}(0 < c,\ \frac{1}{5} \le \xi < \frac{1}{2})$, *the following asymptotic normality of* $\widehat{H}(x_0)$ *holds:*

$$\sqrt{nh}\left(\widehat{H}(x_0) - H(x_0)\right) \xrightarrow{d} N(B, V_1),$$

where $B = \lim_{n\to\infty}(nh^5)^{1/2}B_1$,

$$B_1 = \frac{A_{1,2}}{2}\left[\frac{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3}{(1-F)^5}\right](x_0)$$

and

$$V_1 = A_{2,0}H(x_0).$$

Remark 2 If $h = o(n^{-1/5})$, $B = 0$.

The asymptotic normality of the indirect estimator is easily obtained by using the Slutsky's theorem.

The direct estimator is superior in the sense of the asymptotic variance, but the bias is large. We will consider the bias reduction in Sect. 4. We have the following higher-order asymptotic bias.

Theorem 3 *Let us assume that* (i′) $f(\cdot)$ *is six-times differentiable at* x_0, $f^{(6)}(x_0)$ *is bounded,* (ii′) K *is symmetric and the support is given by a closed interval,* (iii) $K^{(3)}$ *is bounded, and* (iv′) $A_{1,6}$ *is bounded. Then, the higher-order asymptotic bias of* $\widehat{H}(x_0)$ *is*

$$E\left[\widehat{H}(x_0) - H(x_0)\right] = h^2B_1(x_0) + h^4B_2(x_0) + O(h^6 + n^{-1}),$$

where

$$B_2(x_0) = \frac{A_{1,4}}{24}\left[\frac{-60m^2(m')^2m''' + 15m^3m''m''' + 11m^3m'm^{(4)} - m^4m^{(5)}}{m^9} + \frac{210m(m')^3m'' - 73m^2m'(m'')^2 - 105(m')^5}{m^9}\right](x_0).$$

3 Comparison of kernel hazard estimators

Here, we investigate the $AMSE$ of the direct $\widehat{H}(x_0)$ and indirect $\widetilde{H}(x_0)$ in certain special cases. We show that the new estimator $\widehat{H}(x_0)$ performs asymptotically better when $F(\cdot)$ is an exponential or gamma distribution.

Here, we will suppose that $F(\cdot)$ is an exponential, uniform, gamma, Weibull, or beta distribution. The cumulative distribution function of the exponential distribution $Exp(1/\lambda)$ is $F(x) = 1 - exp(-\lambda x)$, and the hazard ratio is constant; that is,

$H(x) = \lambda$. This is one of the most common models of survival analysis. When $F(\cdot)$ is exponential, the $AMSE$s are given by

$$AMSE\left[\widehat{H}(x_0)\right] = \frac{\lambda}{nh} A_{2,0}$$
$$AMSE\left[\widetilde{H}(x_0)\right] = \frac{h^4}{4} A_{1,2}^2 \lambda^6 + \frac{\lambda}{nh} \exp(\lambda x_0) A_{2,0}.$$

Since the squared bias is positive, $AMSE$ of the new estimator is asymptotically smaller regardless of the parameter λ and the point x_0 if $h = o(n^{-1/9})$.

Next, let us assume that $F(\cdot)$ is a uniform distribution $[F(x) = x/b \ (0 < x < b)]$. The hazard ratio in this case is $H(x) = (b-x)^{-1}$. The hazard ratio increases drastically in the tail area of this model. The above $AMSE$s are given by

$$AMSE\left[\widehat{H}(x_0)\right] = \frac{h^4}{4} A_{1,2}^2 \frac{9b^4}{(b-x_0)^{10}} + \frac{1}{nh} \frac{1}{b-x_0} A_{2,0}$$
$$AMSE\left[\widetilde{H}(x_0)\right] = \frac{1}{nh} \frac{b}{(b-x_0)^2} A_{2,0}.$$

We find that the asymptotic bias of $\widetilde{H}(x_0)$ vanishes and the variance of $\widehat{H}(x_0)$ decreases. Their asymptotic performance depends on x_0 and b, but the $AMSE$ of the new $\widehat{H}(x_0)$ is smaller when the life span b is large.

Lastly, let us suppose that F is a gamma $\Gamma(p, 100)$, Weibull $W(q, 100)$, or beta distribution $(100 \times B(r, s))$, where p, q, r and s are their shape parameters. Their scales $(\sigma = 100)$ are moderate. $\Gamma(p, \sigma)$ is the distribution of the sum of $p(\in \mathbb{N})$ $i.i.d.$ random variables of $Exp(\sigma)$; hence, it is one of most important cases. Its asymptotic squared bias, variance, and $AMSE$ for some fixed points x_0 are listed in Table 1, where we have omitted terms in powers of h. $\widehat{H}$ and $\widetilde{H}$ represent those values of $\widehat{H}(x_0)$ and $\widetilde{H}(x_0)$, and every x_0 is each εth quantile of $\Gamma(p, 100)$. The kernel is an Epanechnikov one with $A_{1,2} = 1/5$, $A_{2,0} = 3/10$, and $h = n^{-1/5}$. The coefficients $n^{-4/5}$ have been omitted.

The Weibull distribution $W(q, \sigma)$ is also important in survival analysis because the hazard ratio is proportional to the polynomial degree $(q-1)$; that is, $H(x) = q\sigma^q x^{q-1}$. $W(1, \sigma)$ is the exponential distribution. The beta distribution is often used to describe a distribution whose support is finite, and it has plentiful shapes. Tables 2, 3 and 4 give the least $AMSE$ values using h^* or h^{**} (in Sect. 2.2), where $\widehat{H}$ and $\widetilde{H}$ stand for the $AMSE$ values of $\widehat{H}(x_0)$ and $\widetilde{H}(x_0)$. Every x_0 is each εth quantile of $\Gamma(p, 100)$, $W(q, 100)$, or $(100 \times B(r, s))$. The tables demonstrate that the proposed estimator $\widehat{H}$ performs asymptotically better in most cases of the gamma $\Gamma(p, 100)$. Moreover, the asymptotic performance of our estimator in the Weibull distribution cases is good and comparable to that of the beta cases.

4 Bias reduction and simulation study

As discussed in Sect. 3, the direct estimator performs well; in particular, it has a small variance. If we could reduce the bias, we can get a better estimator. As we can see

Table 1 Asymptotic Bias^2, Var and $AMSE$ with $h = n^{-1/5}$ in gamma case

	Bias^2	Var	$AMSE$	Bias^2	Var	$AMSE$
	$p = 1/2, \varepsilon = 0.05$			$p = 1/2, \varepsilon = 0.1$		
$\widehat{H}$	7.22×10^{-2}	4.01×10^{-2}	0.112	8.24×10^{-5}	2.11×10^{-2}	2.11×10^{-2}
$\widetilde{H}$	6.76×10^{-2}	4.22×10^{-2}	0.110	7.24×10^{-5}	2.33×10^{-2}	2.34×10^{-2}
	$p = 1/2, \varepsilon = 0.25$			$p = 1/2, \varepsilon = 0.5$		
$\widehat{H}$	1.33×10^{-8}	9.52×10^{-3}	9.52×10^{-3}	2.36×10^{-11}	5.65×10^{-3}	5.65×10^{-3}
$\widetilde{H}$	9.78×10^{-9}	1.27×10^{-2}	1.27×10^{-2}	1.41×10^{-11}	1.13×10^{-2}	1.13×10^{-2}
	$p = 1/2, \varepsilon = 0.75$			$p = 1/2, \varepsilon = 0.9$		
$\widehat{H}$	5.35×10^{-13}	4.29×10^{-3}	4.29×10^{-3}	1.69×10^{-15}	3.76×10^{-3}	3.76×10^{-3}
$\widetilde{H}$	3.66×10^{-13}	1.72×10^{-2}	1.72×10^{-2}	7.26×10^{-14}	3.76×10^{-2}	3.76×10^{-2}
	$p = 1/2, \varepsilon = 0.95$			$p = 1/2, \varepsilon = 0.975$		
$\widehat{H}$	1.01×10^{-12}	3.58×10^{-3}	3.58×10^{-3}	1.47×10^{-11}	3.46×10^{-3}	3.46×10^{-3}
$\widetilde{H}$	4.23×10^{-14}	7.16×10^{-2}	7.16×10^{-2}	3.07×10^{-14}	0.139	0.139
	$p = 10, \varepsilon = 0.05$			$p = 10, \varepsilon = 0.1$		
$\widehat{H}$	2.49×10^{-18}	1.56×10^{-4}	1.56×10^{-4}	2.16×10^{-18}	2.55×10^{-4}	2.55×10^{-4}
$\widetilde{H}$	4.46×10^{-19}	1.64×10^{-4}	1.64×10^{-4}	7.87×10^{-20}	2.83×10^{-4}	2.83×10^{-4}
	$p = 10, \varepsilon = 0.25$			$p = 10, \varepsilon = 0.5$		
$\widehat{H}$	2.61×10^{-18}	4.77×10^{-4}	4.77×10^{-4}	1.37×10^{-17}	7.72×10^{-4}	7.72×10^{-4}
$\widetilde{H}$	3.86×10^{-18}	6.36×10^{-4}	6.36×10^{-4}	5.54×10^{-18}	1.54×10^{-3}	1.54×10^{-3}
	$p = 10, \varepsilon = 0.75$			$p = 10, \varepsilon = 0.9$		
$\widehat{H}$	2.84×10^{-16}	1.07×10^{-3}	1.07×10^{-3}	1.19×10^{-14}	1.32×10^{-3}	1.32×10^{-3}
$\widetilde{H}$	1.64×10^{-20}	4.28×10^{-3}	4.28×10^{-3}	1.55×10^{-17}	1.32×10^{-2}	1.32×10^{-2}
	$p = 10, \varepsilon = 0.95$			$p = 10, \varepsilon = 0.975$		
$\widehat{H}$	1.84×10^{-13}	1.45×10^{-3}	1.45×10^{-3}	2.75×10^{-12}	1.56×10^{-3}	1.56×10^{-3}
$\widetilde{H}$	4.98×10^{-17}	2.90×10^{-2}	2.90×10^{-2}	1.501×10^{-16}	6.24×10^{-2}	6.24×10^{-2}

Table 2 *AMSE* values with $h = h^*$ or h^{**} in gamma case

$p = 1/2$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	7.44×10^{-2}	1.14×10^{-2}	1.06×10^{-3}	1.97×10^{-4}
$\widetilde{H}$	7.65×10^{-2}	1.21×10^{-2}	1.25×10^{-3}	3.09×10^{-4}
$p = 1/2$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	7.40×10^{-5}	2.11×10^{-5}	7.26×10^{-5}	1.21×10^{-4}
$\widetilde{H}$	2.08×10^{-4}	2.82×10^{-4}	4.23×10^{-4}	6.73×10^{-4}
$p = 10$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	4.48×10^{-7}	6.45×10^{-7}	1.11×10^{-6}	2.26×10^{-6}
$\widetilde{H}$	3.31×10^{-7}	3.61×10^{-7}	1.50×10^{-6}	3.29×10^{-6}
$p = 10$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	5.39×10^{-6}	1.34×10^{-5}	2.51×10^{-5}	4.57×10^{-5}
$\widetilde{H}$	2.32×10^{-6}	2.25×10^{-5}	5.34×10^{-5}	1.13×10^{-4}

Table 3 *AMSE* values $h = h^*$ or h^{**} in Weibull case

$q = 1/2$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	4.11×10^{-2}	5.68×10^{-3}	3.85×10^{-4}	4.25×10^{-5}
$\widetilde{H}$	4.23×10^{-2}	6.01×10^{-3}	4.47×10^{-4}	6.05×10^{-5}
$q = 1/2$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	9.22×10^{-6}	3.31×10^{-6}	3.59×10^{-7}	2.67×10^{-6}
$\widetilde{H}$	1.84×10^{-5}	1.14×10^{-5}	1.08×10^{-5}	1.17×10^{-5}
$q = 10$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	1.20×10^{-4}	2.65×10^{-4}	9.00×10^{-4}	3.40×10^{-3}
$\widetilde{H}$	1.08×10^{-4}	2.08×10^{-4}	2.11×10^{-4}	2.58×10^{-3}
$q = 10$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	1.38×10^{-2}	5.49×10^{-2}	0.135	0.309
$\widetilde{H}$	1.09×10^{-2}	1.89×10^{-2}	9.97×10^{-2}	0.319

(1) in Sect. 2.2, the asymptotic bias of the proposed hazard estimator $\widehat{H}(\cdot)$ comes from its numerator $\widehat{f_{M_n(X)}}(\cdot)$. We need to apply bias reduction methods to the density estimator $\widehat{f_{M_n(X)}}(\cdot)$.

In this section, we discuss some bias reduction methods. The bias term is complicated, but if we use a 4th order kernel, we have $A_{1,2} = 0$. Thus, we can reduce the convergence order of the bias from $O(h^2)$ to $O(h^4)$. A simple way to construct the

Table 4 $AMSE$ values $h = h^*$ or h^{**} in beta case

$r = 1/2, s = 1/2$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	7.61×10^{-3}	1.20×10^{-3}	1.42×10^{-4}	1.46×10^{-4}
$\widetilde{H}$	7.82×10^{-3}	1.27×10^{-3}	1.61×10^{-4}	1.11×10^{-4}
$r = 1/2, s = 1/2$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	2.35×10^{-3}	0.167	4.57	127
$\widetilde{H}$	1.45×10^{-3}	0.103	2.82	78.3
$r = 2, s = 5$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	1.55×10^{-4}	2.70×10^{-4}	5.35×10^{-4}	1.15×10^{-3}
$\widetilde{H}$	2.15×10^{-4}	2.68×10^{-4}	3.49×10^{-4}	3.34×10^{-4}
$r = 2, s = 5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	3.09×10^{-3}	1.01×10^{-2}	2.41×10^{-2}	5.70×10^{-2}
$\widetilde{H}$	2.14×10^{-3}	8.82×10^{-3}	2.30×10^{-2}	5.74×10^{-2}
$r = 5, s = 2$	$\varepsilon = 0.05$	$\varepsilon = 0.1$	$\varepsilon = 0.25$	$\varepsilon = 0.5$
$\widehat{H}$	7.18×10^{-5}	1.38×10^{-4}	4.20×10^{-4}	1.72×10^{-3}
$\widetilde{H}$	6.36×10^{-5}	1.09×10^{-4}	2.38×10^{-4}	3.34×10^{-4}
$r = 5, s = 2$	$\varepsilon = 0.75$	$\varepsilon = 0.9$	$\varepsilon = 0.95$	$\varepsilon = 0.975$
$\widehat{H}$	9.66×10^{-3}	6.68×10^{-2}	0.261	0.981
$\widetilde{H}$	3.14×10^{-3}	2.17×10^{-2}	7.76×10^{-2}	0.260

4th order kernel was proposed by Jones and Signorini (1997); however, the estimator takes a negative value in some cases.

By applying Ruppert and Cline (1994)'s transformation ($X \mapsto \widehat{F}(X)$) to the denominator $\widehat{f}(\cdot)$ instead of the proposed transformation ($X \mapsto M_n(X)$), the following hazard estimator $\widehat{H}_{RC}$ is derived (see Sect. 2):

$$\widehat{H}_{RC}(x_0) = \frac{\widehat{f}_{\widehat{F}(X)}(x_0)}{1 - F_n(x_0)}.$$

The MSE of the density estimator $\widehat{f}_{\widehat{F}(X)}(x_0)$ was obtained by Ruppert and Cline (1994), and so the following optimal convergence rate of MSE of $\widehat{H}_{RC}(x_0)$ is given by using (1):

$$E\left[\left(\widehat{H}_{RC}(x_0) - H(x_0)\right)^2\right] = O(n^{-8/9}),$$

where $h = O(n^{-1/9})$.

Nielsen (1998) applied a multiple bias correction in hazard estimation, which was proposed by Jones et al. (1995) in kernel density estimation. The hazard estimator in this setting is given by

$$\widetilde{H}_N(x_0) = \frac{\widetilde{H}(x_0)}{nh} \sum_{i=1}^{n} \frac{1}{\widetilde{H}(X_i)} K\left(\frac{x_0 - X_i}{h}\right) I(X_i \geq x_0),$$

where $I(\cdot)$ is the indicator function $I(A) = 1$ (if A occurs), $= 0$ (if A fails). The MSE and asymptotic properties were given by Nielsen (1998). By applying Nielsen (1998)'s bias reduction method to the proposed estimator $\widehat{H}(\cdot)$, we have

$$\widehat{H}_N(x_0) = \frac{\widehat{H}(x_0)}{nh} \sum_{i=1}^{n} \frac{1}{\widehat{H}(X_i)} K\left(\frac{M_n(x_0) - M_n(X_i)}{h}\right).$$

The following MSE is obtained by a similar argument as Jones et al. (1995).

Theorem 4 *Suppose that Theorem 1 holds. Then, the MSE is given by*

$$\begin{aligned} E\left[\widehat{H}_N(x_0) - H(x_0)\right]^2 &= h^8 H^2(x_0)\left[\frac{C_1''}{2m^3}(x_0) - \frac{C_1 m''}{2m^4}(x_0) - 3\frac{C_1' m'}{2m^4}(x_0) + 3\frac{C_1 (m')^2}{2m^5}(x_0)\right]^2 \\ &\quad + \frac{H(x_0)}{nh}\int \{2K(u) - K * K(u)\}^2 du + o\left(h^8 + \frac{1}{nh}\right), \end{aligned}$$

where $C_1(x_0) = B_1(x_0)/H(x_0)$.

As we see from the theorem, the asymptotic order of the variance is the same as the original $\widehat{H}$. The order of the optimal bandwidth is $n^{-1/9}$, and the optimal convergence rate is $n^{-8/9}$.

Terrell and Scott (1980)'s method reduces the asymptotic bias of the kernel density estimator $\widehat{f}(\cdot)$ without losing the nonnegativity property. Many researchers have since applied the method to kernel smoothed function estimators. Hirukawa and Sakudo (2014) applied Terrell and Scott (1980)'s method to asymmetric kernel density estimation for nonnegative data. Funke and Kawka (2015) discussed Terrell and Scott (1980)'s bias reduction method in multivariate asymmetric kernel density estimation.

McCune and McCune (1987) proposed the following modified naive hazard estimator:

$$\widetilde{H}^{\dagger}(x_0) = \frac{\widehat{f}^{\dagger}(x_0)}{1 - F_n(x_0)} = \frac{\{\widehat{f}_h(x_0)\}^{4/3}\{\widehat{f}_{2h}(x_0)\}^{-1/3}}{1 - F_n(x_0)},$$

where $\widehat{f}_h(\cdot)$ and $\widehat{f}_{2h}(\cdot)$ are kernel density estimators with bandwidth parameters h and $2h$, respectively. We propose the following modified direct hazard estimators:

$$\widehat{H}^{\dagger}(x_0) = \{\widehat{H}_h(x_0)\}^{\frac{4}{3}}\{\widehat{H}_{2h}(x_0)\}^{-\frac{1}{3}} = \frac{\widehat{f}^{\dagger}_{M_n(X)}(x_0)}{1 - F_n(x_0)},$$

Table 5 MSE values with LCV bandwidths in $Exp(1)$ case

$n = 50$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	5.63×10^{-2}	2.26×10^{-2}	4.34×10^{-2}	0.125
(sd)	5.34×10^{-2}	2.10×10^{-2}	2.16×10^{-2}	4.14×10^{-2}
$\widehat{H}_{RC}$	7.83×10^{-2}	4.95×10^{-2}	8.97×10^{-2}	0.279
(sd)	0.106	8.31×10^{-2}	0.146	0.427
$\widehat{H}_N$	6.44×10^{-2}	$\underline{1.83 \times 10^{-2}}$	4.37×10^{-2}	0.222
(sd)	6.45×10^{-2}	8.69×10^{-3}	1.56×10^{-2}	4.18×10^{-2}
$\widehat{H}^{\dagger}$	$\underline{4.29 \times 10^{-2}}$	2.14×10^{-2}	$\underline{2.55 \times 10^{-2}}$	$\underline{0.107}$
(sd)	4.90×10^{-2}	1.30×10^{-2}	1.43×10^{-2}	4.16×10^{-2}
$\widetilde{H}$	7.32×10^{-2}	5.11×10^{-2}	8.37×10^{-2}	0.357
(sd)	6.53×10^{-2}	5.23×10^{-2}	0.229	1.07
$n = 200$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	4.31×10^{-2}	1.90×10^{-2}	4.07×10^{-2}	0.111
(sd)	2.54×10^{-2}	7.60×10^{-3}	9.76×10^{-3}	2.04×10^{-2}
$\widehat{H}_{RC}$	3.27×10^{-2}	$\underline{1.29 \times 10^{-2}}$	4.30×10^{-2}	0.151
(sd)	3.54×10^{-2}	1.70×10^{-2}	3.70×10^{-2}	8.63×10^{-2}
$\widehat{H}_N$	5.01×10^{-2}	2.35×10^{-2}	4.53×10^{-2}	0.228
(sd)	3.21×10^{-2}	1.23×10^{-3}	7.80×10^{-3}	1.92×10^{-2}
$\widehat{H}^{\dagger}$	$\underline{2.60 \times 10^{-2}}$	2.50×10^{-2}	$\underline{2.08 \times 10^{-2}}$	0.110
(sd)	2.20×10^{-2}	5.02×10^{-3}	7.28×10^{-3}	2.46×10^{-2}
$\widetilde{H}$	5.98×10^{-2}	3.70×10^{-2}	3.04×10^{-2}	$\underline{5.33 \times 10^{-2}}$
(sd)	3.34×10^{-2}	2.59×10^{-2}	2.91×10^{-2}	0.115

where $\widehat{f}^{\dagger}_{M_n(X)}(\cdot)$ denotes the following bias-corrected transformed density estimator by Terrell and Scott (1980)'s method:

$$\widehat{f}^{\dagger}_{M_n(X)}(x_0) = \left\{\widehat{f}_{M_n(X),h}(x_0)\right\}^{4/3}\left\{\widehat{f}_{M_n(X),2h}(x_0)\right\}^{-1/3},$$

where $\widehat{f}^{\dagger}_{M_n(X),h}(\cdot)$ and $\widehat{f}^{\dagger}_{M_n(X),2h}(\cdot)$ are transformed density estimators with bandwidth parameters h and $2h$, respectively.

By using Terrell and Scott (1980)'s result, we can get the asymptotic bias of $\widehat{H}^{\dagger}(x_0)$:

$$E[\widehat{H}^{\dagger}(x_0) - H(x_0)] = \frac{2B_1^2(x_0) - 4B_2(x_0)H(x_0)}{H(x_0)}h^4,$$

Table 6 *MSE* values with *LCV* bandwidths in $\Gamma(1/2, 1)$ case

$n = 50$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	0.762	8.66×10^{-2}	3.44×10^{-2}	9.22×10^{-2}
(sd)	0.675	9.77×10^{-2}	1.96×10^{-2}	3.73×10^{-2}
$\widehat{H}_{RC}$	1.04	0.259	0.196	0.630
(sd)	1.77	0.512	0.770	2.12
$\widehat{H}_N$	0.748	4.89×10^{-2}	3.08×10^{-2}	0.157
(sd)	0.690	7.82×10^{-2}	1.08×10^{-2}	3.53×10^{-2}
$\widehat{H}^{\dagger}$	0.882	5.82×10^{-2}	3.59×10^{-2}	8.02×10^{-2}
(sd)	0.725	7.99×10^{-2}	1.27×10^{-2}	2.28×10^{-2}
$\widetilde{H}$	0.704	0.183	0.178	0.553
(sd)	0.725	0.203	0.615	1.66
$n = 200$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	0.750	7.58×10^{-2}	3.24×10^{-2}	0.111
(sd)	0.343	4.36×10^{-2}	8.19×10^{-3}	1.68×10^{-2}
$\widehat{H}_{RC}$	0.293	5.70×10^{-2}	3.78×10^{-2}	9.19×10^{-2}
(sd)	0.358	9.02×10^{-2}	5.59×10^{-2}	0.158
$\widehat{H}_N$	0.688	2.47×10^{-2}	3.38×10^{-2}	0.162
(sd)	0.344	2.34×10^{-2}	6.59×10^{-3}	1.90×10^{-2}
$\widehat{H}^{\dagger}$	0.815	3.89×10^{-2}	4.09×10^{-2}	7.72×10^{-2}
(sd)	0.363	2.68×10^{-2}	$5,73 \times 10^{-3}$	1.14×10^{-2}
$\widetilde{H}$	0.596	0.114	4.76×10^{-2}	8.35×10^{-2}
(sd)	0.383	9.13×10^{-2}	5.46×10^{-2}	0.221

where $B_1(x_0)$ and $B_2(x_0)$ are as given in Theorem 3. On the other hand, the variance is given by

$$\begin{aligned} V[\widehat{H}^{\dagger}(x_0)] &= V\left[\frac{4}{3}\widehat{H}_h(x_0) - \frac{1}{3}\widehat{H}_{2h}(x_0)\right] + O(n^{-1}) \\ &= \frac{16}{9}V\left[\widehat{H}_h(x_0)\right] + \frac{1}{9}V\left[\widehat{H}_{2h}(x_0)\right] - \frac{8}{9}Cov\left[\widehat{H}_h(x_0), \widehat{H}_{2h}(x_0)\right] \\ &\quad + O(n^{-1}). \end{aligned}$$

Then, we have the following asymptotic variance and *MSE* of $\widehat{H}^{\dagger}(x_0)$.

Theorem 5 *Suppose that Theorem 1 holds. Then, the asymptotic variance of* $\widehat{H}^{\dagger}(x_0)$ *is given by*

$$V[\widehat{H}^{\dagger}(x_0)] = \frac{1}{nh}H(x_0)\int \left\{2K^2(u) + K(2u)K(u)\right\} du + O\left(\frac{1}{nh^{1/2}}\right),$$

Table 7 MSE values with LCV bandwidths in $\Gamma(10, 1)$ case

$n = 50$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	5.36×10^{-4}	1.95×10^{-3}	$\underline{8.09 \times 10^{-3}}$	$\underline{2.57 \times 10^{-2}}$
(sd)	3.00×10^{-4}	1.13×10^{-3}	3.50×10^{-3}	8.69×10^{-3}
$\widehat{H}_{RC}$	4.01×10^{-3}	7.96×10^{-3}	2.08×10^{-2}	6.17×10^{-2}
(sd)	4.62×10^{-3}	1.08×10^{-2}	2.64×10^{-2}	9.50×10^{-2}
$\widehat{H}_N$	5.61×10^{-4}	$\underline{1.45 \times 10^{-3}}$	8.14×10^{-3}	2.86×10^{-2}
(sd)	3.24×10^{-4}	7.38×10^{-4}	3.78×10^{-3}	1.03×10^{-2}
$\widehat{H}^\dagger$	$\underline{4.37 \times 10^{-4}}$	1.61×10^{-3}	8.16×10^{-3}	2.86×10^{-2}
(sd)	1.92×10^{-4}	8.17×10^{-4}	3.49×10^{-3}	9.35×10^{-3}
$\widetilde{H}$	1.37×10^{-3}	3.33×10^{-3}	1.02×10^{-2}	8.00×10^{-2}
(sd)	1.14×10^{-3}	3.15×10^{-3}	4.17×10^{-2}	0.270
$n = 200$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	6.74×10^{-4}	1.92×10^{-3}	7.88×10^{-3}	2.33×10^{-2}
(sd)	1.32×10^{-4}	4.84×10^{-4}	1.67×10^{-3}	4.35×10^{-3}
$\widehat{H}_{RC}$	1.86×10^{-3}	2.62×10^{-3}	1.00×10^{-2}	3.52×10^{-2}
(sd)	1.91×10^{-3}	3.19×10^{-3}	1.01×10^{-2}	2.62×10^{-2}
$\widehat{H}_N$	6.25×10^{-4}	$\underline{1.53 \times 10^{-3}}$	8.06×10^{-3}	2.45×10^{-2}
(sd)	1.31×10^{-4}	3.32×10^{-4}	1.97×10^{-3}	5.71×10^{-3}
$\widehat{H}^\dagger$	$\underline{4.72 \times 10^{-4}}$	1.73×10^{-3}	8.26×10^{-3}	2.49×10^{-2}
(sd)	8.67×10^{-5}	3.84×10^{-4}	1.74×10^{-3}	4.80×10^{-3}
$\widetilde{H}$	1.03×10^{-3}	2.13×10^{-3}	$\underline{3.63 \times 10^{-3}}$	$\underline{1.05 \times 10^{-2}}$
(sd)	5.70×10^{-4}	1.51×10^{-3}	3.47×10^{-3}	2.69×10^{-2}

and so, the MSE is as follows:

$$E\left[\widehat{H}^\dagger(x_0) - H(x_0)\right]^2$$
$$= \frac{2B_1^2(x_0) - 4B_2(x_0)H(x_0)}{H(x_0)}h^8$$
$$+ \frac{1}{nh}H(x_0)\int\left\{2K^2(u) + K(2u)K(u)\right\}du + O\left(h^{10} + \frac{1}{nh^{1/2}}\right).$$

As we see from Theorem 5, the asymptotic bias is reduced, and the optimal convergence rate is $n^{-8/9}$.

In practice, to find optimal bandwidth parameters of the proposed estimators is a important problem. By estimating their $AMSE$ which depend on unknown functions, we can obtain theoretically optimal bandwidth parameters (plug-in method). However, the nonparametric estimators of the $AMSE$ of the proposed estimator require us to choose more bandwidth parameters. In this paper, we consider so-called cross-

Table 8 MSE values with LCV bandwidths in $W(1/2, 1)$ case

$n = 50$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	0.191	1.60×10^{-2}	1.69×10^{-3}	7.52×10^{-4}
(sd)	0.197	1.65×10^{-2}	1.27×10^{-3}	3.06×10^{-4}
$\widehat{H}_{RC}$	0.337	3.86×10^{-2}	2.83×10^{-2}	2.53×10^{-2}
(sd)	0.583	7.92×10^{-2}	1.84×10^{-2}	8.60×10^{-3}
$\widehat{H}_N$	0.176	1.27×10^{-2}	1.33×10^{-3}	1.03×10^{-3}
(sd)	0.192	1.47×10^{-2}	6.10×10^{-4}	3.10×10^{-4}
$\widehat{H}^{\dagger}$	0.199	1.45×10^{-2}	1.05×10^{-3}	4.61×10^{-4}
(sd)	0.204	1.65×10^{-2}	8.92×10^{-4}	1.95×10^{-4}
$\widetilde{H}$	0.199	2.66×10^{-2}	1.17×10^{-2}	4.50×10^{-2}
(sd)	0.222	3.38×10^{-2}	3.79×10^{-2}	0.158
$n = 200$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	0.183	1.56×10^{-2}	1.89×10^{-3}	6.55×10^{-4}
(sd)	0.101	8.33×10^{-3}	4.63×10^{-4}	1.36×10^{-4}
$\widehat{H}_{RC}$	9.16×10^{-2}	8.32×10^{-3}	2.49×10^{-2}	2.69×10^{-2}
(sd)	0.114	1.33×10^{-2}	9.77×10^{-3}	3.92×10^{-3}
$\widehat{H}_N$	0.152	1.07×10^{-2}	1.68×10^{-3}	9.74×10^{-4}
(sd)	9.54×10^{-2}	7.07×10^{-3}	3.11×10^{-4}	1.64×10^{-4}
$\widehat{H}^{\dagger}$	0.183	1.23×10^{-2}	1.11×10^{-3}	3.63×10^{-4}
(sd)	0.104	8.17×10^{-3}	2.90×10^{-4}	8.36×10^{-5}
$\widetilde{H}$	0.151	1.47×10^{-2}	2.74×10^{-3}	5.76×10^{-3}
(sd)	0.112	1.32×10^{-2}	3.46×10^{-3}	1.50×10^{-2}

validation method instead. Although the cross-validation method may also be effective in the direct hazard estimation as Patil (1993) showed in the naive hazard function estimation, the $IMSE$ may sometimes diverge. Therefore, we consider a local cross-validation method, which is introduced Vieu (1991a) in kernel smoothed regression estimation. For a hazard estimator $\bar{H}_h(\cdot)$, it is given by

$$\arg\min_{h>0} \int \{\bar{H}_h(x) - H(x)\}^2 w_n(x)\mathrm{d}x,$$

where w_n is a (local) weight function. Let us define $i_n(x_0) := [x_0 - d_n, x_0 + d_n]$ where the interval is the minimum which includes the closest realization value to x_0 (say $\acute{X}$). If we choose $w_n(x) = I(x \in i_n(x_0))$, it holds for large enough n that

$$\arg\min_{h>0} \int \left\{\bar{H}_h(x) - H(x)\right\}^2 I(x \in i_n(x_0))\mathrm{d}x$$
$$\approx \arg\min_{h>0} \left(d_n \bar{H}_h^2(x_0) - \frac{\bar{H}_h(\acute{X})}{n} \frac{1}{1 - F_n(\acute{X})} \right).$$

Table 9 MSE values with LCV bandwidths in $W(10, 1)$ case

$n = 50$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	0.137	1.58	11.2	<u>53.7</u>
(sd)	5.78×10^{-2}	0.784	5.57	24.0
$\widehat{H}_{RC}$	0.218	1.55	11.1	172
(sd)	0.318	2.33	28.7	592
$\widehat{H}_N$	0.100	<u>1.25</u>	<u>10.8</u>	54.6
(sd)	6.35×10^{-2}	0.730	6.25	28.4
$\widehat{H}^\dagger$	<u>7.81×10^{-2}</u>	1.35	11.4	58.3
(sd)	3.85×10^{-2}	0.731	5.80	25.7
$\widetilde{H}$	0.271	2.43	15.3	145
(sd)	0.241	2.16	26.0	398
$n = 200$	$\varepsilon = 0.25$	$\varepsilon = 0.5$	$\varepsilon = 0.75$	$\varepsilon = 0.9$
$\widehat{H}$	0.152	1.76	11.0	46.7
(sd)	2.84×10^{-2}	0.394	2.77	12.1
$\widehat{H}_{RC}$	<u>6.31×10^{-2}</u>	<u>0.518</u>	<u>3.43</u>	<u>22.7</u>
(sd)	8.33×10^{-2}	0.621	3.71	43.8
$\widehat{H}_N$	0.101	1.20	10.0	43.1
(sd)	2.67×10^{-2}	0.354	3.20	15.2
$\widehat{H}^\dagger$	8.32×10^{-2}	1.36	11.2	48.8
(sd)	1.88×10^{-2}	0.364	2.94	13.1
$\widetilde{H}$	0.198	1.65	8.14	26.1
(sd)	0.120	1.06	5.97	48.1

The approximate minimization problem does not include integration fortunately, and so locally cross-validated (LCV) bandwidth is much more easy to obtain.

Next, we compare proposed direct hazard estimators $\widehat{H}(\cdot)$, $\widehat{H}_{RC}(\cdot)$, $\widehat{H}_N(\cdot)$, $\widehat{H}^\dagger(\cdot)$ and $\widetilde{H}(\cdot)$ by simulation study. We simulated MSE values of the hazard estimators 100,000 times. Tables 5, 6, 7, 8 and 9 shows each average, and we drew underlines the minimum MSE values for each cases. All kernel function were the Epanechnikov one. The bandwidth parameters were chosen by the local cross-validation and obtained by simulation in advance. We first see that MSE values of both modified estimators $\widehat{H}_N(\cdot)$ and $\widehat{H}^\dagger(\cdot)$ are smaller than those of $\widehat{H}(\cdot)$ especially when $n = 50$. In $W(10, 1)$ cases, $\widehat{H}_{RC}(\cdot)$ especially seems to be quite precise as n increases, though the estimated values are not stable as 'sd' (standard deviation of the MSE) values show. 'sd' also shows us that the variance of the direct estimators are smaller than the naive $\widetilde{H}(\cdot)$, and so the proposed estimators are numerically stable.

To sum up, we recommend $\widehat{H}_N(\cdot)$ and $\widehat{H}^\dagger(\cdot)$ if the underlying distribution F seems to be gamma or exponential. In Weibull ($W(p, 1)$) cases, $\widehat{H}_N(\cdot)$ and $\widehat{H}^\dagger(\cdot)$ are still recommended, but we recommend $\widehat{H}_{RC}(\cdot)$ if both n and p is large enough.

Acknowledgements The authors would like to appreciate the editor's and referees' valuable comments that helped us to improve this manuscript significantly. The authors gratefully acknowledge *JSPS KAKENHI Grant Number JP15K11995 and JP16H02790*.

Appendices: Proofs of Theorems

Proof of Theorem 1 For simplicity, we will use the following notation,

$$\eta(z) = \int_{-\infty}^{z} F(u)\mathrm{d}u, \qquad \gamma(z) = \int_{-\infty}^{z} \eta(u)\mathrm{d}u,$$
$$M(z) = z - \eta(z) \quad \text{and} \quad m(z) = M'(z) = 1 - F(z).$$

To begin with, we consider the following stochastic expansion of the direct estimator:

$$\begin{aligned}
&\widehat{H}(x_0)\\
&= \frac{1}{h}\int K\left(\frac{M(w)-M(x_0)}{h}\right)\mathrm{d}F_n(w)\\
&\quad + \frac{1}{h^2}\int K'\left(\frac{M(w)-M(x_0)}{h}\right)\Big\{[\eta(w)-\eta_n(w)]-[\eta(x_0)-\eta_n(x_0)]\Big\}\mathrm{d}F_n(w)\\
&\quad + \frac{1}{h^3}\int K''\left(\frac{M(w)-M(x_0)}{h}\right)\Big\{[\eta(w)-\eta_n(w)]-[\eta(x_0)-\eta_n(x_0)]\Big\}^2\mathrm{d}F_n(w)\\
&\quad + \frac{1}{h^4}\int K^{(3)}\left(\frac{M_n^*(w)-M(x_0)}{h}\right)\Big\{[\eta(w)-\eta_n(w)]-[\eta(x_0)-\eta_n(x_0)]\Big\}^3\mathrm{d}F_n(w)\\
&= J_1 + J_2 + J_3 + J_4^* \qquad \text{(say)},
\end{aligned}$$

where

$$\eta_n(w) = \int_{-\infty}^{w} F_n(u)\mathrm{d}u = \frac{1}{n}\sum_{i=1}^{n}(w - X_i)_+$$

and $M_n^*(w)$ is a r.v. between $M_n(w)$ and $M(w)$ with probability 1.

The main term of the expectation of $\widehat{H}(x_0)$ is given by J_1, as we will show the later. Since J_1 is a sum of $i.i.d.$ random variables, the expectation can be obtained directly:

$$\begin{aligned}
E[J_1] &= E\left[\frac{1}{h}\int K\left(\frac{M(w)-M(x_0)}{h}\right)\mathrm{d}F_n(w)\right]\\
&= \frac{1}{h}\int K\left(\frac{M(w)-M(x_0)}{h}\right) f(w)\mathrm{d}w\\
&= \int K(u)H(M^{-1}(M(x_0)+hu))\mathrm{d}u\\
&= H(x_0) + \frac{h^2}{2}A_{1,2}\left[\frac{(1-F)\{(1-F)f'' + 4ff'\} + 3f^3}{(1-F)^5}\right](x_0) + O(h^4).
\end{aligned}$$

Combining the following second moment,

$$\begin{aligned}
&\frac{1}{nh^2}\int K^2\left(\frac{M(w)-M(x_0)}{h}\right)f(w)\mathrm{d}w\\
&\quad=\frac{1}{nh}\int K^2(u)H(M^{-1}(M(x_0)+hu))\mathrm{d}u\\
&\quad=\frac{A_{2,0}}{nh}H(x_0)+O(n^{-1}),
\end{aligned}$$

we get the variance,

$$V[J_1]=\frac{1}{nh}H(x_0)A_{2,0}+O(n^{-1}).$$

Next, we consider the following representation of J_2

$$J_2=\frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}K'\left(\frac{M(X_i)-M(x_0)}{h}\right)Q(X_i,X_j),$$

where

$$Q(x_i,x_j)=[\eta(x_i)-(x_i-x_j)_+]-[\eta(x_0)-(x_0-x_j)_+].$$

Using the conditional expectation, we get the following equation:

$$\begin{aligned}
E[J_2]&=\frac{1}{nh^2}\sum_{j=1}^{n}E\left[K'\left(\frac{M(X_i)-M(x_0)}{h}\right)Q(X_i,X_j)\right]\\
&=\frac{1}{nh^2}E\left[K'\left(\frac{M(X_i)-M(x_0)}{h}\right)E\left[\sum_{j=1}^{n}Q(X_i,X_j)\,\middle|\,X_i\right]\right]\\
&=\frac{1}{nh^2}E\left[K'\left(\frac{M(X_i)-M(x_0)}{h}\right)\left\{\eta(X_i)-\left[\eta(x_0)-(x_0-X_i)_+\right]\right\}\right]\\
&=\frac{1}{nh}\int K'(u)\left\{\eta(M^{-1}(M(x_0)+hu))-\eta(x_0)+(x_0-M^{-1}(M(x_0)+hu))_+\right\}\\
&\qquad\times H(M^{-1}(M(x_0)+hu))\mathrm{d}u\\
&=\frac{1}{nh}\int K'(u)\,O(hu)H(x_0)\mathrm{d}u=O\left(\frac{1}{n}\right).
\end{aligned}$$

Next, we have

$$
\begin{aligned}
J_2^2 &= \frac{1}{n^4h^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n} K'\left(\frac{M(X_i)-M(x_0)}{h}\right)K'\left(\frac{M(X_s)-M(x_0)}{h}\right)\\
&\quad \times Q(X_i,X_j)Q(X_s,X_t)\\
&= \frac{1}{n^4h^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{n}\sum_{t=1}^{n} \Xi(i,j,s,t) \qquad \text{(say)}.
\end{aligned}
$$

After taking the conditional expectation, we find that if all of the (i, j, s, t) are different,

$$
E[\Xi(i,j,s,t)] = E\left[E\left\{\Xi(i,j,s,t)|X_i,X_s\right\}\right] = 0,
$$

and

$$
\begin{aligned}
&E[\Xi(i,j,s,t)] = 0 \quad (\text{if } i=j \text{ and all of } (i,s,t) \text{ are different}),\\
&E[\Xi(i,j,s,t)] = 0 \quad (\text{if } i=s \text{ and all of } (i,j,t) \text{ are different}),\\
&E[\Xi(i,j,s,t)] = 0 \quad (\text{if } i=t \text{ and all of } (i,j,s) \text{ are different}),
\end{aligned}
$$

the term in which $j = t$ and all of the (i, j, s) are different is the main term of $E[J_2^2]$. If $j = t$ and all of the (i, j, s) are different, we have

$$
\begin{aligned}
&E[\Xi(i,j,s,t)]\\
&\quad = \frac{n(n-1)(n-2)}{n^4h^4}E\Bigg[K'\left(\frac{M(X_i)-M(x_0)}{h}\right)K'\left(\frac{M(X_s)-M(x_0)}{h}\right)\\
&\qquad\qquad \times Q(X_i,X_j)Q(X_s,X_j)\Bigg].
\end{aligned}
$$

Using the conditional expectation of $Q(X_i, X_j)Q(X_s, X_j)$ given X_i and X_s, we find that

$$
\begin{aligned}
&E\Big[E\Big\{Q(X_i,X_j)Q(X_s,X_j)\ \Big|\ X_i,X_s\Big\}\Big]\\
&= E\Big[\eta(X_i)\eta(x_0)+\eta(X_s)\eta(x_0)-\eta^2(x_0)+2\gamma(x_0)-\eta(X_i)\eta(X_s)\\
&\quad -(x+X_i-2\min(x,X_i))\eta(\min(x,X_i))-2\gamma(\min(x,X_i))\\
&\quad -(x+X_s-2\min(x,X_s))\eta(\min(x,X_s))-2\gamma(\min(x,X_s))\\
&\quad +(X_i+X_s-2\min(X_i,X_s))\eta(\min(X_i,X_s))+2\gamma(\min(X_i,X_s))\Big].
\end{aligned}
$$

Therefore, the entire expectation of the last row is

$$\begin{aligned}
&E\Bigg[K'\left(\frac{M(X_i)-M(x_0)}{h}\right)K'\left(\frac{M(X_s)-M(x_0)}{h}\right)\\
&\quad\times(X_i+X_s-2\min(X_i,X_s))\eta(\min(X_i,X_s))+2\gamma(\min(X_i,X_s))\Bigg]\\
&=\int\Bigg[\int_{\infty}^{w}K'\left(\frac{M(z)-M(x_0)}{h}\right)K'\left(\frac{M(w)-M(x_0)}{h}\right)\\
&\quad\times\Big\{(-z+w)\eta(z)+2\gamma(z)\Big\}f(w)\mathrm{d}z\\
&\quad+\int_{w}^{\infty}K'\left(\frac{M(z)-M(x_0)}{h}\right)K'\left(\frac{M(w)-M(x_0)}{h}\right)\\
&\quad\times\{(z-w)\eta(w)+2\gamma(w)\}f(z)dz\Bigg]f(w)\mathrm{d}w.
\end{aligned}$$

Finally, we get

$$\begin{aligned}
&E\Bigg[K'\left(\frac{M(X_i)-M(x_0)}{h}\right)K'\left(\frac{M(X_s)-M(x_0)}{h}\right)\\
&\quad\times\{X_i+X_s-2\min(X_i,X_s)\}\eta(\min(X_i,X_s))+2\gamma(\min(X_i,X_s))\Bigg]\\
&=-h^2\int K'\left(\frac{M(w)-M(x_0)}{h}\right)f(w)\mathrm{d}w\\
&\quad\times\Bigg(\Bigg[W\left(\frac{M(w)-M(x_0)}{h}\right)\left(\{\eta(x_0)+(-x_0+w)F(x_0)\}\left[\frac{f}{m^2}\right](x_0)\right.\\
&\quad\left.+\{(-x_0+w)\eta(x_0)+2\gamma(x_0)\}\left[\frac{f'm-fm'}{m^3}\right](x_0)\right)+O(h)\Bigg]\\
&\quad+\Bigg[\left(1-W\left(\frac{M(w)-M(x_0)}{h}\right)\right)\\
&\quad\times\left(\eta(w)\frac{f}{m^2}(x_0)+\{(x_0-w)\eta(w)+2\gamma(w)\}\left[\frac{f'm-fm'}{m^3}\right](x_0)\right)+O(h)\Bigg]\Bigg)\\
&=h^4\left[\frac{Ff^2}{m^4}(x_0)+\left\{\frac{f'm-fm'}{m^3}\left(2\eta\frac{f}{m^2}+2\gamma\frac{f'm-fm'}{m^3}\right)\right](x_0)\right\}+O(h^5).
\end{aligned}$$

After similar calculations of the other terms, we find that if $j=t$ and all of the (i,j,s) are different,

$$E[\Xi(i,j,s,t)]=O\left(\frac{1}{n}\right).$$

In addition, it is easy to see that the expectations of the other combinations of (i,j,s,t) are $o(n^{-1})$. As a result, we have

$$E[J_2^2] = O(n^{-1}) \quad \text{and} \quad V[J_2] = O(n^{-1}).$$

The moments of J_3 can be obtained in a similar manner; we find that

$$E[J_3] = O(n^{-1}) \quad \text{and} \quad V[J_3] = O(n^{-2}).$$

Finally, we will obtain upper bounds of $|E[J_4^*]|$ and $E[(J_4^*)^2]$. From the assumption of Theorem 1, we can see

$$\begin{aligned}
&\left| E\left[\frac{1}{h^4} K^{(3)}\left(\frac{M_n^*(X_i) - M(x_0)}{h}\right) \{[\eta(X_i) - \eta_n(X_i)] - [\eta(x_0) - \eta_n(x_0)]\}^3\right]\right| \\
&\leq \frac{\max_u |K^{(3)}(u)|}{h^4} E\left[\{[\eta(X_i) - \eta_n(X_i)] - [\eta(x_0) - \eta_n(x_0)]\}^3\right] \\
&= O\left(\frac{1}{n^2h^4}\right).
\end{aligned}$$

Similarly, it follows that

$$\begin{aligned}
&E[(J_4^*)^2] \\
&= \frac{1}{h^8} E\left[\left(\int K^{(3)}\left(\frac{M_n^*(w) - M(x_0)}{h}\right) \{[\eta(w) - \eta_n(w)] - [\eta(x_0) - \eta_n(x_0)]\}^3 \mathrm{d}F_n(w)\right)^2\right] \\
&\leq \frac{\max_u (K^{(3)}(u))^2}{n^2h^8} \sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\ \{[\eta(X_i) - \eta_n(X_i)] - [\eta(x_0) - \eta_n(x_0)]\}^3\right. \\
&\quad \left. \times \left\{[\eta(X_j) - \eta_n(X_j)] - [\eta(x_0) - \eta_n(x_0)]\right\}^3\ \right] \\
&= O\left(\frac{1}{n^4h^8}\right).
\end{aligned}$$

To sum up, we conclude that $J_2 + J_3 + J_4^*$ is asymptotically negligible for fixed x_0. The main bias of $\widehat{H}(x_0)$ comes from J_1. From the Cauchy–Schwarz inequality, we find that the main term of the variance coincides with $V[J_1]$. Now, we can get the $AMSE$ of the direct estimator and prove Theorem 1. □

Proof of Theorems 2 and 3 It follows from the above discussion that

$$\begin{aligned}
&\sqrt{nh}\left\{\widehat{H}(x_0) - H(x_0)\right\} \\
&= \sqrt{nh}\big\{J_1 - H(x_0)\big\} + o_P(1) \\
&= (\sqrt{nh}h^2)B_1 + \sqrt{nh}\left\{J_1 - H(x_0) - h^2B_1\right\} + o_P(1).
\end{aligned}$$

Since J_1 is a sum of $i.i.d.$ random variables and the expectation of the second term is $o(1)$, asymptotic normality holds for Theorem 2. □

Furthermore, we can show that

$$
\begin{aligned}
E\left[\widehat{H}(x_0)\right] &= E[J_1] + O(n^{-1}) \\
&= \int K(u)H(M^{-1}(M(x_0)+hu))\mathrm{d}u + O(n^{-1}),
\end{aligned}
$$

and we can directly get Theorem 3 by taking a Taylor expansion. □

Proof of Theorem 4 We follow the proof of Jones et al. (1995). $\widehat{H}_N(x_0)$ is written as

$$
\widehat{H}_N(x_0) = \widehat{H}(x_0)\widehat{\alpha}(x_0) = H(x_0)\left\{1 + \frac{\widehat{H}(x_0) - H(x_0)}{H(x_0)}\right\}\{1 + (\widehat{\alpha}(x_0) - 1)\}.
$$

It follows from the asymptotic expansion that

$$
\begin{aligned}
\widehat{\alpha}(x_0) \approx \frac{1}{n}\sum_{i=1}^{n} \frac{1}{hH(X_i)} K\left(\frac{M_n(x_0) - M_n(X_i)}{h}\right) \\
\times \left[1 - \frac{\widehat{H}(X_i) - H(X_i)}{H(X_i)} + \left\{\frac{\widehat{H}(X_i) - H(X_i)}{H(X_i)}\right\}^2\right].
\end{aligned}
$$

By taking the expectation of the ith term in this sum conditional on X_i, we have

$$
\begin{aligned}
&E\left[\frac{1}{hH(X_i)} K\left(\frac{M_n(x_0) - M_n(X_i)}{h}\right)\right. \\
&\quad\left.\times \left[1 - \frac{\widehat{H}(X_i) - H(X_i)}{H(X_i)} + \left\{\frac{\widehat{H}(X_i) - H(X_i)}{H(X_i)}\right\}^2\right] \middle| X_i\right] \\
&= \frac{1}{hH(X_i)} K\left(\frac{M(x_0) - M(X_i)}{h}\right)\left[1 - \frac{h^2 B_1(X_i) + h^4 B_2(X_i)}{H(X_i)} + \left(\frac{h^2 B_1(X_i)}{H(X_i)}\right)^2\right] \\
&\quad + O_P((nh)^{-1}) + o_P(h^4).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
E[\widehat{\alpha}(x_0)] &= 1 - h^2 \frac{B_1}{H}(M^{-1}(M(x_0) + hu)) \\
&\quad + h^4\left[-\frac{B_2(x_0)}{H(x_0)} + \left(\frac{B_1(x_0)}{H(x_0)}\right)^2\right] + O((nh)^{-1}) + o(h^4) \\
&= 1 - h^2 C_1(x_0) \\
&\quad + h^4\left[-\frac{B_2(x_0)}{H(x_0)} + \left(\frac{B_1(x_0)}{H(x_0)}\right)^2 + \frac{C_1''}{2m^3}(x_0) - \frac{C_1 m''}{2m^4}(x_0)\right. \\
&\quad \left. - 3\frac{C_1' m'}{2m^4}(x_0) + 3\frac{C_1 (m')^2}{2m^5}(x_0)\right] + O((nh)^{-1}) + o(h^4).
\end{aligned}
$$

It follows that

$$\begin{aligned}
&E[\widehat{H}_N(x_0) - H(x_0)] \\
&= h^2 B_1(x_0) + h^4[B_2(x_0) - B_1(x_0)C_1(x_0)] - h^2 H(x_0)C_1(x_0) \\
&\quad + h^4 H(x_0)\left[-\frac{B_2(x_0)}{H(x_0)} + \left(\frac{B_1(x_0)}{H(x_0)}\right)^2 + \frac{C_1''}{2m^3}(x_0) - \frac{C_1 m''}{2m^4}(x_0) \right. \\
&\quad \left. - 3\frac{C_1' m'}{2m^4}(x_0) + 3\frac{C_1(m')^2}{2m^5}(x_0) \right] + O((nh)^{-1}) + o(h^4) \\
&= h^4 H(x_0)\left[\frac{C_1''}{2m^3}(x_0) - \frac{C_1 m''}{2m^4}(x_0) - 3\frac{C_1' m'}{2m^4}(x_0) + 3\frac{C_1(m')^2}{2m^5}(x_0) \right] \\
&\quad + O((nh)^{-1}) + o(h^4).
\end{aligned}$$

From the proof of Theorem 1, we can see that the following approximation holds:

$$\begin{aligned}
\widehat{H}_N(x_0) &= \frac{1}{n^2 h}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{M_n(x_0) - M_n(X_i)}{h}\right)\frac{K\left(\frac{M_n(x_0)-M_n(X_j)}{h}\right)}{\sum_{\ell=1}^n K\left(\frac{M_n(X_j)-M_n(X_\ell)}{h}\right)} \\
&\approx \frac{1}{n^2 h}\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{M(x_0) - M(X_i)}{h}\right)\frac{K\left(\frac{M(x_0)-M(X_j)}{h}\right)}{\sum_{\ell=1}^n K\left(\frac{M(X_j)-M(X_\ell)}{h}\right)} \\
&=: H_N(x_0)
\end{aligned}$$

$H_N(x_0)$ can be seen as a Jones et al. (1995)'s density estimate for the random variable $M(X)$ at the point $M(x_0)$. Then, the asymptotic variance is given by

$$V[\widehat{H}_N(x_0)] \approx \frac{H(x_0)}{nh}\int \left\{2K(u) - K * K(u)\right\}^2 \mathrm{d}u$$

(see Jones et al. 1995). □

Proof of Theorem 5 To obtain the asymptotic variance of the modified estimator $\widehat{H}^\dagger$, we need to calculate the covariance term $Cov[\widehat{H}_h(x_0), \widehat{H}_{2h}(x_0)]$ as shown in Sect. 4. From the proof of Theorem 1, it is easy to see

$$\begin{aligned}
&Cov\left[\widehat{H}_h(x_0), \widehat{H}_{2h}(x_0)\right] \\
&= E\left[\widehat{H}_h(x_0)\widehat{H}_{2h}(x_0)\right] - E\left[\widehat{H}_h(x_0)\right]E\left[\widehat{H}_{2h}(x_0)\right] \\
&= \frac{1}{2n^2h^2}E\left[\sum_{i=1}^n\sum_{j=1}^n K\left(\frac{M(X_i) - M(x_0)}{h}\right)K\left(\frac{M(X_j) - M(x_0)}{2h}\right)\right] \\
&\quad - \left\{H^2(x_0) + 5h^2 H(x_0)B_1(x_0) + O(h^4 + n^{-1})\right\} + O\left(\frac{1}{nh^{1/2}}\right).
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
&Cov\left[\widehat{H}_h(x_0), \widehat{H}_{2h}(x_0)\right] \\
&= \frac{1}{2nh^2} E\left[K\left(\frac{M(X_1)-M(x_0)}{h}\right) K\left(\frac{M(X_1)-M(x_0)}{2h}\right)\right] \\
&\quad -\frac{1}{n}\left\{H^2(x_0) + 5h^2 H(x_0) B_1(x_0) + O(h^4 + n^{-1})\right\} + O\left(\frac{1}{nh^{1/2}}\right) \\
&= \frac{1}{nh} H(x_0) \int K(2u)K(u)\mathrm{d}u + O\left(\frac{1}{nh^{1/2}}\right)
\end{aligned}$$

and the desired result. □

References

Chen, S. M., Hsu, Y. S., Liaw, J. T. (2009). On kernel estimators of density ratio. *Statistics*, *43*(5), 463–479.

Ćwik, J., Mielniczuk, J. (1989). Estimating density ratio with application to discriminant analysis. *Communications in Statistics-Theory and Methods*, *18*(8), 3057–3069.

De Haan, L., Ferreira, A. (2007). *Extreme value theory*: *An introduction*. New York, NY: Springer.

Funke, B., Kawka, R. (2015). Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. *Computational Statistics & Data Analysis*, *92*, 148–162.

Gumbel, E. (1958). *Statistics of extremes*. New York, NY: Columbia University Press.

Hirukawa, M., Sakudo, M. (2014). Nonnegative bias reduction methods for density estimation using asymmetric kernels. *Computational Statistics & Data Analysis*, *75*, 112–123.

Jones, M., Signorini, D. (1997). A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association*, *92*(439), 1063–1073.

Jones, M. C., Linton, O., Nielsen, J. P. (1995). A simple bias reduction method for density estimation. *Biometrika*, *82*(2), 327.

McCune, E., McCune, S. (1987). On improving convergence rates for nonnegative kernel failure-rate function estimators. *Statistics & Probability Letters*, *6*(2), 71–76.

Müller, H. G., Wang, J. L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, *50*(1), 61.

Murthy, V. (1965). Estimation of jumps, reliability and hazard rate. *The Annals of Mathematical Statistics*, *36*(3), 1032–1040.

Nielsen, J. P. (1998). Multiplicative bias correction in kernel hazard estimation. *Scandinavian Journal of Statistics*, *25*(3), 541–553.

Patil, P. (1993). On the least squares cross-validation bandwidth in hazard rate estimation. *The Annals of Statistics*, *21*(4), 1792–1810.

Quintela-del Río, A. (2007). Plug-in bandwidth selection in kernel hazard estimation from dependent data. *Computational Statistics & Data Analysis*, *51*(12), 5800–5812.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, *27*(3), 832–837.

Ruppert, D., Cline, D. B. H. (1994). Bias reduction in kernel density estimation by smoothed empirical transformations. *The Annals of Statistics*, *22*(1), 185–210.

Tanner, M. A., Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *The Annals of Statistics*, *11*(3), 989–993.

Terrell, G. R., Scott, D. W. (1980). On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, *8*(5), 1160–1163.

Vieu, P. (1991a). Nonparametric regression: Optimal local bandwidth choice. *Journal of the Royal Statistical Society Series B (Methodological)*, *53*(2), 453–464.

Vieu, P. (1991b). Quadratic errors for nonparametric estimates under dependence. *Journal of Multivariate Analysis*, *39*(2), 324–347.

Watson, G., Leadbetter, M. (1964). Hazard analysis ii. *Sankhyā*: *The Indian Journal of Statistics, Series A* *26*(1), 101–116.