VARIABLE BANDWIDTH NONPARAMETRIC HAZARD RATE ESTIMATION

DIMITRIOS BAGKAVOS AND PRAKASH PATIL

Abstract. A smoothing parameter inversely proportional to the square root of the true density is known to produce kernel estimates of the density having faster bias rate of convergence. We show that in the case of kernel-based nonparametric hazard rate estimation, a smoothing parameter inversely proportional to the square root of the true hazard rate leads to a mean square error rate of order $n^{-8/9}$, an improvement over the standard second order kernel. An adaptive version of such a procedure is considered and analyzed.

1. Introduction.

For a given random sample $X_1, X_2, ..., X_n$ of independent failure times with common probability density function $f(x)$ and distribution function $F(x)$, the most commonly used kernel-based estimator of the hazard rate, $\lambda(x)(= f(x)/(1-F(x))$, for $F(x) < 1)$, is

$$\hat{\lambda}(x|h) = h^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_{(i)}}{h}\right)(n-i+1)^{-1} \tag{1}$$

where K is any second order positive kernel, h denotes the bandwidth and $X_{(i)}$ is the i-th order statistic. Since Watson and Leadbetter (1964) introduced the above estimator, several of its asymptotic properties have been studied in the literature. For example, its mean square error (MSE) properties for complete samples are given in Rice and Rosenblatt (1976). Besides mean square error, Tanner and Wong (1983) proved the asymptotic normality of the above estimator for censored data using the Hajék projection method, while Lo, Mack and Wang (1989) achieved the same result via strong representation of the Kaplan-Meier estimator. Smoothing parameter selection by least squares cross-validation is investigated by Sarda and Vieu (1991) and Patil (1993). González-Manteiga, Cao and Marron (1996) have employed the bootstrap for bandwidth selection.

If a second order kernel and an asymptotically optimal bandwidth are used, then the mean square error rate of convergence of $\hat{\lambda}(x|h)$ is $n^{-4/5}$. It would be possible to achieve a faster rate by using higher order kernels. However this leads to the undesirable feature of having hazard rate estimates that could take negatives values. To alleviate this problem, in analogy to the proposal of Hall and Marron (1988) for kernel density estimation, we allow the bandwidth in $\hat{\lambda}(x|h)$ to be inversely proportional to the square root of the true hazard rate itself. It should be noted that Hall, Hu and Marron (1995) have provided the fine tuning of the proposal in Hall and Marron (1988). Since in our case the effective interval over which hazard rate is estimated is finite, the issues addressed in Hall, Hu and Marron (1995) do not play a role. Although an estimator, obtained by allowing the bandwidth in $\hat{\lambda}(x|h)$ to be

Key words and phrases. Hazard rate estimation, kernel, mean integrated square error, smoothness, variable bandwidth.

inversely proportional to the square root of the true hazard rate, is not useful in practice, it is still helpful to know that, as shown in Section 2, this estimator achieves a faster rate convergence without taking negative values.

To construct a practically feasible estimator, we replace the unknown hazard rate in the bandwidth function by its usual kernel estimate. In Section 3 it is shown that such an estimator is almost as good as the ideal estimator as far as the rate of convergence to the true curve is concerned. In Section 4, for practical implementation, a graphical bandwidth selection method is presented with illustrations of the estimate for simulated and real life data.

It should be mentioned that Silverman (1986) has considered variable bandwidth for hazard rate estimation, albeit indirectly and without the theoretical analysis considered here. In contrast to our direct approach, the hazard rate there is estimated as a ratio of density and survival function estimators, with the density estimate being based on variable bandwidth. The results obtained here for the estimator $\hat{\lambda}(x|h)$ could be derived for an estimator obtained based on the approach described in Silverman (1986) and have been discussed in Bagkavos (2003). Another attempt to study variable smoothing in hazard rate estimation is by Tanner (1983). Noting the difficulty in establishing theoretical properties of a hazard rate estimator that uses variable smoothing, Tanner (1983) proves the strong consistency of such an estimator with a deterministic form of variable smoothing. Thus the contribution that this article makes is to provide asymptotic formulae for bias and variance when the variable bandwidth has a deterministic form. Further, it establishes the convergence rate of an (adaptive) variable bandwidth estimator when the variable bandwidth is stochastic and provides comparison between the adaptive and the ideal estimators. It also provides a practically useful method to implement the adaptive estimator.

2. Variable bandwidth estimator and mean square error analysis.

We modify estimator $\hat{\lambda}(x|h)$ defined in (1) by taking bandwidth proportional to $\lambda^{-\frac{1}{2}}$ to define a new estimator that makes use of variable bandwidth,

$$\tilde{\lambda}_n(x|h) = \frac{1}{h}\sum_{i=1}^{n} \frac{\lambda(X_{(i)})^{\frac{1}{2}} K\left(\frac{x-X_{(i)}}{h}\lambda(X_{(i)})^{\frac{1}{2}}\right)}{n-i+1}. \tag{2}$$

This estimator is a natural extension of the density estimator defined in Hall and Marron (1988) to the case of hazard rate estimation.

Note that the bandwidth $h(x) \equiv h\lambda(x)^{-\frac{1}{2}}$, used in the definition of $\tilde{\lambda}_n(x|h)$ involves the true hazard function. This makes the above estimator infeasible in practice. However, in the next section, we show that this *ideal* variable bandwidth estimator has better bias and hence MSE (since their variances are of the same order) properties than its fixed bandwidth counterpart.

Unless stated otherwise, we assume standard kernel conditions, that is, the kernel function $K : \mathbb{R} \to \mathbb{R}$, satisfies:

A.1
$$\int K(z)\, dz = 1.$$

A.2 K is nonnegative, symmetric and five times differentiable.

A.3 K vanishes outside a compact interval.

We note here two implications of conditions A.2 and A.3. First, both the fourth moment of the kernel and its supremum are finite. Second, the first two derivatives of the kernel are bounded. Typically, estimation of the hazard function is done in $[0,T]$, where $T = \sup\{x|F(x) < 1-\varepsilon\}$, for some $\varepsilon > 0$.

The asymptotic mean and variance formulae of $\tilde{\lambda}_n(x|h)$ are summarized in the following theorem.

Theorem 2.1. *Under assumptions A1,A3 for the kernel, and assuming that λ has four continuous derivatives and that it is bounded away from zero on $(0,T)$,*

$$\mathbb{E}\left\{\tilde{\lambda}_n(x|h)\right\} = \lambda(x) + g(x)h^4 \int z^4 K(z)\,dz + o(h^4)$$

$$\mathbb{V}\mathrm{ar}\left\{\tilde{\lambda}_n(x|h)\right\} = \frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\frac{1}{nh}\int K^2(z)\,dz + o\left((nh)^{-1}\right)$$

uniformly in $x \in (0,T)$ as $h \to 0$, $n \to +\infty$ and $nh \to +\infty$, with

$$g(x) = \frac{1}{24\lambda(x)^5}\Big(24\lambda'(x)^4 - 36\lambda'(x)^2\lambda''(x)^2\lambda(x) + 6\lambda''(x)^2\lambda^2(x) \\ + 8\lambda'(x)\lambda'''(x)\lambda^2(x) - \lambda^{(4)}(x)\lambda^3(x)\Big).$$

Proof. First we obtain an integral expression for the bias, which is then evaluated by applying a transformation and finally its terms are expanded as power series of the transformed variable. To begin, from [18] we have

$$\mathbb{E}\left\{\hat{\lambda}(x|h)\right\} = \int \lambda(u)K_h(u-x)[1-F^n(u)]\,du.$$

Then,

$$\begin{aligned}
\mathbb{E}\left\{\tilde{\lambda}_n(x|h)\right\} &= \mathbb{E}\left\{h^{-1}\sum_{i=1}^{n}\frac{\lambda(X_{(i)})^{\frac{1}{2}}K\left(\frac{x-X_{(i)}}{h}\lambda(X_{(i)})^{\frac{1}{2}}\right)}{n-i+1}\right\} \\
&= h^{-1}\int [1-F^n(u)]\lambda(u)^{1/2}K\left(\frac{x-u}{h}\lambda(u)^{1/2}\right)\lambda(u)\,du \\
&= h^{-1}\int \lambda(u)^{\frac{3}{2}}K\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right)du + o\left(n^{-1}\right)
\end{aligned}$$

because $|F(x)| < 1$ and for large n, $|F(x)|^n = o(n^{-1})$. Now, applying the transformation $x-u = hz$ yields

$$\begin{aligned}
\mathbb{E}\left\{\tilde{\lambda}_n(x|h)\right\} &= \int \lambda(x-hz)^{3/2}K\left(z\lambda(x-hz)^{1/2}\right)dz + o\left(n^{-1}\right) \\
&= \int \lambda(x)^{\frac{3}{2}}\frac{\lambda(x-hz)^{\frac{3}{2}}}{\lambda(x)^{\frac{3}{2}}}K\left(z\lambda(x)^{\frac{1}{2}}\frac{\lambda(x-hz)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}}\right)dz + o\left(n^{-1}\right).
\end{aligned}$$

Now set

$$u(z) = \frac{\lambda(x-z)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} \tag{3}$$

and

$$z\lambda(x)^{\frac{1}{2}} = y \Leftrightarrow dz = \lambda(x)^{-\frac{1}{2}}dy. \tag{4}$$

Then we have

$$u\left(\frac{y}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x - \frac{y}{\lambda^{\frac{1}{2}}(x)}\right)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} \Leftrightarrow u\left(\frac{hy}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x - \frac{hy}{\lambda(x)^{\frac{1}{2}}}\right)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}}. \tag{5}$$

Further set

$$\eta = \frac{h}{\lambda(x)^{1/2}} \tag{6}$$

and substitute back in (5) to get

$$\mathbb{E}\left\{\tilde{\lambda}_n(x|h)\right\} = \lambda(x)\int u^3(\eta y)K\Big(yu(\eta y)\Big)\,dy + o\left(n^{-1}\right). \tag{7}$$

In order to analyze this mean expression further we expand $u(\eta y)$ in Taylor series around 0.

$$u(\eta y) = u(0) + \eta y u'(0) + \frac{(\eta y)^2}{2}u''(0) + \frac{(\eta y)^3}{3!}u'''(0) + \frac{(\eta y)^4}{4!}u^{(4)}(0) + o(\eta^5). \tag{8}$$

Raising to the third power and rearranging,

$$\begin{aligned} u^3(\eta y) = u^3(0) &+ \eta[3yu^2(0)u'(0)] + \eta^2\left[3y^2u(0)u'(0)^2 + \frac{3}{2}y^2u^2(0)u''(0)\right] \\ &+\eta^3\left[y^3u'(0)^3 + \frac{3}{3!}y^3u^2(0)u'''(0) + 3y^3u'(0)u''(0)\right] \\ &+\eta^4\Big[y^4u^2(0)u'(0)u'''(0) + \frac{3}{2}y^4u'(0)^2u''(0) \\ &+ \frac{3}{4}y^4u(0)u''(0)^2 + \frac{1}{8}y^4u''''(0)\Big] + o(\eta^5). \end{aligned} \tag{9}$$

From (3), $u(0) = 1$. Set

$$\begin{aligned} \rho(y) &= yu(\eta y) \\ &= y + \eta y^2u'(0) + \frac{\eta^2y^3}{2}u''(0) + \frac{\eta^3y^4}{3!}u'''(0) + \frac{\eta^4y^5}{4!}u^{(4)}(0) + o(\eta^5). \end{aligned}$$

Also,

$$K\Big(yu(\eta y)\Big) =$$

$$K(y) + \rho(y)K'(y) + \rho^2(y)\frac{K''(y)}{2} + \rho^3(y)\frac{K'''(y)}{3!} + \rho^4(y)\frac{K''''(y)}{4!} + o(\rho^5(y)). \tag{10}$$

Rearranging (10) we get

$$
\begin{gathered}
K\Big(yu(\eta y)\Big) = \\
K(y) + \eta\{y^2u'(0)\}K'(y) + \eta^2\left\{\frac{y^3}{2}u''(0)K'(y) + \frac{y^4u'(0)^2}{2}K''(y)\right\} + \\
\eta^3\left\{\frac{y^4}{3!}u'''(0)K'(y) + \frac{y^5}{2}u'(0)u''(0)K''(y) + \frac{y^6}{3!}u'(0)^3K'''(y)\right\} + \\
\eta^4\left\{\frac{y^5}{4!}u''''(0)K'(y) + \frac{y^6}{8}u''(0)^2K''(y) + \frac{y^6}{6}u'(0)u^{(3)}(0)K''(y)\right\}^4 + o(\eta^5).
\end{gathered}
$$

Multiplying (9) with (10), noting that $u(0) = 1$ gives

$$
\begin{gathered}
K(z) + \eta\left\{3zu'(0)K(z) + z^2u'(0)K'(z)\right\} + \\
\eta^2\Bigg\{K(z)\left(3z^2u'(0)^2 + \frac{3}{2}z^2u''(0)\right) + \\
K'(z)\left(\frac{z^3}{2}u''(0) + 3z^3u'(0)^2\right) + \frac{z^4}{2}u'(0)^2K''(z)\Bigg\} + \\
\eta^3\Bigg\{K(z)\left(z^3u'(0)^3 + \frac{3}{3!}z^3u'''(0)\right) + K'(z)\Bigg(\frac{z^4}{3!}u'''(0) + \frac{3z^4}{2}u'(0)u''(0) + \\
3z^4u'(0)^3 + \frac{3}{2}z^4u'(0)u'''(0)\Bigg) + K''(z)\left(\frac{z^5}{2}u'(0)u''(0) + \frac{3z^5}{2}u'(0)^3\right) + \\
K'''(z)\frac{z^6}{3}u'(0)^3\Bigg\} + \\
\eta^4\Bigg\{u''(0)\frac{3}{4}z^4K(z) + u'(0)^2u''(0)\Big(\frac{3}{2}z^4K(z) + \frac{9}{2}z^5K'(z) + \frac{9}{4}z^6K''(z) \\
+ \frac{1}{4}z^7K'''(z)\Big) + u''(0)^2\left(\frac{z^6K''(z)}{8} + \frac{3}{4}z^4K(z) + \frac{3}{4}z^5K'(z)\right) + \\
u'(0)u'''(0)\left(z^4K(z) + z^5K'(z) + \frac{z^6K''(z)}{6}\right) + \\
u'(0)^4\left(\frac{1}{4!}z^8K''''(z)\frac{3}{2}z^6K''(z) + \frac{1}{2}z^7K'''(z) + z^5K'(z)\right) \\
+ u''''(0)\left(\frac{1}{8}z^4K(z) + \frac{z^5}{4!}K'(z)\right)\Bigg\}. \qquad (11)
\end{gathered}
$$

We integrate (11) by integrating each one of the coefficients of the powers of the η's. Note that

$$\int yK(y)dy = \int y^2 K'(y)dy = 0$$

because we assume a symmetric kernel. Thus the integral of the coefficient of η is 0. Also the coefficient of η^3 is a sum of products of even and odd functions, thus its integral will be zero. Denote as $l(z)$ the coefficient of η^2 in (11) and note that

$$\int 3y^2 K(y)\,dy + 3\int y^3 K'(y)\,dy + \int \frac{y^4}{2} K''(y)\,dy = y^3 K(y) + \frac{y^4}{2} K'(y) \tag{12}$$

and

$$\int \left\{ \frac{3}{2} y^2 K(y) + \frac{y^3}{2} K'(y) \right\} dy = \frac{y^3}{2} K(y). \tag{13}$$

Rearranging $l(z)$, integrating and using (12) and (13), yields

$$\int l(z)\,dz = u'(0)^2 \left[y^3 K(y) + \frac{y^4}{2} K'(y) \right]_{-\infty}^{+\infty} + u''(0) \left[\frac{y^3}{2} K(y) \right]_{-\infty}^{+\infty} = 0$$

since we assumed that the kernel K vanishes outside a compact interval. Now the coefficient of η^4 can be written as

$$\begin{aligned}
u'(0)^2 u''(0) &\left\{ \frac{3}{2} y^4 K(y) + \frac{9}{2} y^5 K'(y) + \frac{9}{4} y^6 K''(y) + \frac{1}{4} y^7 K'''(y) \right\} + \\
u''(0)^2 &\left\{ \frac{3}{4} y^4 K(y) + \frac{3}{4} y^5 K'(y) + \frac{1}{8} y^6 K''(y) \right\} + \\
u'(0)^4 &\left\{ y^5 K'(y) + \frac{3}{2} y^6 K''(y) + \frac{1}{2} y^7 K'''(y) + \frac{1}{4!} y^8 K''''(y) \right\} + \\
u'(0) u'''(0) &\left\{ y^4 K(y) + y^5 K'(y) + \frac{1}{6} y^6 K''(y) \right\} + \\
u''''(0) &\left\{ \frac{1}{8} y^4 K(y) + \frac{y^5}{4!} K'(y) \right\}.
\end{aligned}$$

By partial integration on each one of the square brackets we get the coefficient of η^4

$$\left(-\frac{1}{12} u''''(0) + u'(0) u'''(0) + 5u'^4(0) + \frac{3}{4} u''(0)^2 - 6u'(0)^2 u''(0) \right) \int y^4 K(y)\,dy.$$

Set the coefficient of the integral as $g_2(x)$. Calculating the derivatives of $u(z)$ and plugging them to the bias formula will give us the final expression for the bias. The derivatives of

$u(z)$ at $z=0$ are

$$
\begin{aligned}
u'(0) &= -\frac{\lambda'(x)}{2\lambda(x)},\\
u''(0) &= \frac{\lambda''(x)}{2\lambda(x)} - \frac{\lambda'(x)}{4\lambda^2(x)},\\
u'''(0) &= -\frac{\lambda'''(x)}{\lambda(x)} + \frac{3\lambda'(x)\lambda''(x)}{4\lambda^2(x)} - \frac{3\lambda'(x)^3}{\lambda^3(x)}\\
u''''(0) &= -\frac{15\lambda'(x)^4}{16\lambda^4(x)} + \frac{3\lambda'(x)\lambda''(x)}{\lambda^2(x)} - \frac{3\lambda''(x)^2}{\lambda^2(x)} - \frac{\lambda'(x)\lambda'''(x)}{\lambda^2(x)} + \frac{\lambda''''(x)}{2\lambda(x)}.
\end{aligned}
$$

Substituting the values of the derivatives back to $g_2(x)$ gives

$$
\begin{gathered}
g_2(x) = \\
\frac{1}{12}\left\{\frac{15\lambda'(x)^4}{16\lambda^4(x)} - \frac{3\lambda'(x)\lambda''(x)}{\lambda^2(x)} + \frac{3\lambda''(x)^2}{\lambda^2(x)} + \frac{\lambda'(x)\lambda'''(x)}{\lambda^2(x)} - \frac{\lambda''''(x)}{2\lambda(x)}\right\}\\
+\left\{-\frac{\lambda'(x)}{2\lambda(x)}\right\}\left\{-\frac{\lambda'''(x)}{\lambda(x)} + \frac{3\lambda'(x)\lambda''(x)}{4\lambda^2(x)} - \frac{3\lambda'(x)^3}{\lambda^3(x)}\right\} + 5\left\{-\frac{\lambda'(x)}{2\lambda(x)}\right\}^4\\
+\frac{3}{4}\left\{\frac{\lambda''(x)}{2\lambda(x)} - \frac{\lambda'(x)}{4\lambda^2(x)}\right\}^2 - 6\left\{-\frac{\lambda'(x)}{2\lambda(x)}\right\}^2\left\{\frac{\lambda''(x)}{2\lambda(x)} - \frac{\lambda'(x)}{4\lambda^2(x)}\right\}.
\end{gathered}
$$

Simplifying, the above expression becomes

$$
\begin{gathered}
g_2(x) = \\
\frac{25}{64}\frac{\lambda'(x)^4}{\lambda^4(x)} - \frac{3}{16}\frac{\lambda'(x)^2\lambda''(x)}{\lambda^3(x)} + \frac{1}{16}\frac{\lambda''(x)^2}{\lambda^2(x)} + \frac{1}{12}\frac{\lambda'(x)\lambda'''(x)}{\lambda^2(x)} - \frac{1}{24}\frac{\lambda''''(x)}{\lambda(x)}\\
-\frac{1}{2}\frac{\lambda'(x)\left(-\frac{3}{8}\frac{\lambda'(x)^3}{\lambda^3(x)} + \frac{3}{4}\frac{\lambda'(x)\lambda''(x)}{\lambda^2(x)} - \frac{1}{2}\frac{\lambda'''(x)}{\lambda(x)}\right)}{\lambda(x)} + \frac{3}{4}\left(-\frac{1}{4}\frac{\lambda'(x)^2}{\lambda^2(x)} + \frac{1}{2}\frac{\lambda''(x)}{\lambda(x)}\right)^2\\
-\frac{3}{2}\frac{\lambda'(x)^2\left(-\frac{1}{4}\frac{\lambda'(x)^2}{\lambda^2(x)} + \frac{1}{2}\frac{\lambda''(x)}{\lambda(x)}\right)}{\lambda^2(x)}.
\end{gathered}
$$

Taking common factors and rearranging we find

$$
\begin{aligned}
g_2(x) = -\frac{-24\lambda'(x)^4 + 36\lambda'(x)^2\lambda''(x)^2\lambda(x) - 6\lambda''(x)^2\lambda^2(x)}{24\lambda^4(x)}&\\
+\frac{-8\lambda'(x)\lambda'''(x)\lambda^2(x) + \lambda''''(x)\lambda^3(x)}{24\lambda^4(x)}&.
\end{aligned}
$$

Recall that $\eta = h/\lambda^{\frac{1}{2}}(x)$. Then,

$$
\eta g_2(x) = \frac{h^4}{\lambda^2(x)} g_2(x) = h^4 g(x) \tag{14}
$$

By (7), the integral of expression (11) and (14), the bias expression of theorem 2.1 is proved. As about the variance, set

$$\mathrm{I}_n(F) = \sum_{i=1}^{n} \frac{1}{n-i+1} \binom{n}{i-1} F(u)^{i-1}(1-F(u))^{n-i+1}.$$

and note that for $0 < F < 1$

$$n\mathrm{I}_n(F) \to \frac{1}{1-F} \quad \text{as} \quad n \to +\infty$$

see for example lemma 6 in [18]. Using this we find

$$\mathbb{E}\left\{\tilde{\lambda}_n(x|h)^2\right\} =$$

$$\frac{1}{h^2}\int \sum_{i=1}^{n} \frac{\lambda(u)K^2\left(\frac{x-u}{h}\lambda^{\frac{1}{2}}(u)\right)}{n-i+1}\binom{n}{i-1}F(u)^{i-1}(1-F(u))^{n-i}f(u)\,du.$$

$$= \frac{1}{nh^2}\int \lambda(u)^2 K^2\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right) n\mathrm{I}_n(F)\,du.$$

Thus, as $n \to +\infty$

$$\mathbb{E}\left\{\tilde{\lambda}_n(x|h)^2\right\} \simeq \frac{1}{nh^2}\int \frac{\lambda(u)^2}{1-F(u)} K^2\left(\frac{x-u}{h}\lambda(u)^{\frac{1}{2}}\right) du.$$

Set $Q(u) = \lambda(u)^2/(1-F(u))$ and $x - u = hz$. Then,

$$\mathbb{E}\left\{\tilde{\lambda}_n(x|h)^2\right\} \simeq \frac{1}{nh}\int Q(x-hz)K^2\left(z\lambda(x-hz)^{\frac{1}{2}}\right) dz$$

$$= \frac{1}{nh}\int \frac{Q(x)}{Q(x)} Q(x-hz) K^2\left(z\frac{\lambda(x)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}}\lambda(x-hz)^{\frac{1}{2}}\right) dz. \tag{15}$$

Now, set

$$\nu(z) = \frac{\lambda(x-z)^2}{1-F(x-z)}\left(\frac{\lambda(x)^2}{1-F(x)}\right)^{-1}.$$

Applying (4) gives

$$\nu\left(\frac{y}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x - y\lambda(x)^{-\frac{1}{2}}\right)^2}{1-F\left(x-y\lambda(x)^{-\frac{1}{2}}\right)}\left(\frac{\lambda(x)^2}{1-F(x)}\right)^{-1}.$$

Applying (6) yields

$$\nu\left(\frac{hy}{\lambda(x)^{\frac{1}{2}}}\right) = \frac{\lambda\left(x - hy\lambda(x)^{-\frac{1}{2}}\right)^2}{1-F\left(x-hy\lambda(x)^{-\frac{1}{2}}\right)}\left(\frac{\lambda(x)^2}{1-F(x)}\right)^{-1}$$

and so

$$\nu(\eta y) = \frac{\lambda(x-\eta y)^2}{1-F(x-\eta y)}\left(\frac{\lambda(x)^2}{1-F(x)}\right)^{-1}. \tag{16}$$

Using (3)-(6) and (16) back in (15) gives

$$
\begin{aligned}
\mathbb{E}\left\{\tilde{\lambda}_n(x|h)^2\right\} &\simeq \frac{1}{nh}\int Q(x)\lambda(x)^{-\frac{1}{2}}\nu(\eta y)K^2(yu(\eta y))\,dy \\
&= \frac{1}{nh}\frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\int \nu(\eta y)K^2(yu(\eta y))\,dy. \qquad (17)
\end{aligned}
$$

Squaring (10) gives the Taylor expansion for K^2. Expand $\nu(\eta y)$ in Taylor series around 0 and substitute both expansions to (17). Then, after some algebra similar to that needed for the bias case, the definition of variance for $\tilde{\lambda}_n(x|h)$ becomes

$$
\mathbb{V}\mathrm{ar}\left\{\tilde{\lambda}_n(x|h)\right\} = \frac{1}{nh}\frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\int K^2(z)\,dz - n^{-1}\left\{\mathbb{E}\tilde{\lambda}_n(x|h)\right\}^2
$$

from which the variance expression of theorem 2.1 follows. □

Remark 2.1. *From Theorem 2.1 the mean square error distance between $\tilde{\lambda}_n(x|h)$ and $\lambda(x)$ can easily shown to be $O(n^{-\frac{8}{9}})$. For this, first note that the values of h (in the bandwidth function, $h\lambda^{\frac{1}{2}}(x)$) which minimize the mean square error of $\tilde{\lambda}_n(x|h)$ are asymptotic to constant multiples of $n^{-\frac{1}{9}}$. Then, for such h's,*

$$
\mathbb{E}\left\{\tilde{\lambda}_n(x|h)-\lambda(x)\right\}^2 = O(n^{-\frac{8}{9}}).
$$

Remark 2.2. *With $h \sim n^{-\frac{1}{9}}$ one could establish that as $n \to +\infty$*

$$
\frac{\sqrt{nh}\left(\tilde{\lambda}_n(x|h)-\mathbb{E}\tilde{\lambda}_n(x|h)\right)}{\sigma(x)} \to N(0,1)
$$

where

$$
\sigma^2(x) = \frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\int K^2(t)\,dt.
$$

Thus the distance between $\tilde{\lambda}_n(x|h)$ and $\lambda(x)$ is $O_p(n^{-\frac{4}{9}})$. For an outline see the proof of the asymptotic normality part of theorem 3.2 given in the appendix.

Remark 2.3. *In the discussion above, the bandwidth of the ideal estimator (i.e. h in $h\lambda^{-1/2}(x)$) is denoted and treated as constant. The case of random bandwidth could be analyzed as follows. Assume that the kernel is compactly supported and twice continuously differentiable and that a random bandwidth $\hat{h}$ satisfies $n^{\frac{1}{9}}\hat{h} \to c$ in probability, where $0 < c < +\infty$. Then, using an argument similar to that of Abramson (1982) for the proof of equation (2) there, one can show that*

$$
\tilde{\lambda}_n(x|\hat{h}) = \tilde{\lambda}_n(x|cn^{-\frac{1}{9}}) + o_p(n^{-\frac{4}{9}})
$$

as $n \to +\infty$ for every $x \in (0,T)$.

3. Adaptive estimation.

In order to get a practically useful estimator we replace $\lambda(x)$ in (2) by a simple kernel estimator. This leads to the definition of the so-called *adaptive* estimator of the hazard rate function,

$$\hat{\lambda}_n(x|h_1,h_2) = \frac{1}{h_2}\sum_{i=1}^{n} \frac{\hat{\lambda}(X_{(i)}|h_1)^{\frac{1}{2}} K\left(\frac{x-X_{(i)}}{h_2}\hat{\lambda}(X_{(i)}|h_1)^{\frac{1}{2}}\right)}{n-i+1} \tag{18}$$

where $\hat{\lambda}(x|h)$ is the 'pilot' estimator defined in (1). Practical choice of the bandwidths involved in (18) is discussed in section 4.

Note that for the estimator based on random bandwidths we will replace h_1, h_2 in the definition of $\hat{\lambda}_n(x|h_1,h_2)$, by $\hat{h}_1$ and $\hat{h}_2$.

To quantify the distance between the ideal and adaptive estimators we prove that up to terms of $o_p(n^{-4/9})$ the adaptive equals the ideal plus a remainder term, and that this term is asymptotically normally distributed. For that first define,

$$\mu(x|h) = \mathbb{E}\hat{\lambda}(x|h), \quad L(x) = K(x) + xK'(x) \text{ and } L_1(z) = zK'(z).$$

Define also,

$$T(x|h_1,h_2) = \frac{1}{2nh_1h_2}\sum_{i=1}^{n} t(X_i, x|h_1,h_2), \tag{19}$$

where

$$t(u,x|h_1,h_2) = \mathbb{E}\Bigg\{\lambda(X_i)^{-\frac{1}{2}}\left(\frac{K\left(\frac{X_i-u}{h_1}\right)}{1-F(X_i)} - h_1\mu(X_i|h_1)\right) \times L\left\{\left(\frac{x-X_i}{h_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}(1-F(X_i))^{-1}\Bigg\}.$$

Then we have the following theorem.

Theorem 3.1. *Assume that the kernel satisfies conditions A1-A3. Suppose that $\lambda > 0$ is three times differentiable with the third derivative satisfying a Lipschitz condition of unit order. Also, we assume that if the bandwidth is random then, with probability 1 as $n \to +\infty$*

$$n^{-a-\frac{1}{5}} < \hat{h}_1 < n^{a-\frac{1}{5}}, \quad \text{where } a < \frac{1}{5}$$
$$\text{and } \eta n^{-\frac{1}{9}} < \hat{h}_2 < \rho n^{-\frac{1}{9}}$$

with $\rho > \eta > 0$. Then

$$\hat{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \tilde{\lambda}_n(x|\hat{h}_2) + T(x|\hat{h}_1,\hat{h}_2) + o_p\left(n^{-\frac{4}{9}}\right). \tag{20}$$

Proof. Observe that

$$\hat{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}\right)}{1-F(X_i)}$$
$$+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}\right)\left\{\frac{1}{1-F_n(X_i)}-\frac{1}{1-F(X_i)}\right\}$$
$$=\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2)+o_p(n^{-\frac{1}{2}})$$

where

$$\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\hat{\lambda}(X_i|\hat{h}_1)^{\frac{1}{2}}\right)}{1-F(X_i)}.$$

Also,

$$\tilde{\lambda}_n(x|\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\lambda(X_i)^{\frac{1}{2}}\right)}{1-F(X_i)}$$
$$+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\lambda(X_i)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\lambda(X_i)^{\frac{1}{2}}\right)\left\{\frac{1}{1-F_n(X_i)}-\frac{1}{1-F(X_i)}\right\}.$$

Hence, we can write

$$\tilde{\lambda}_n(x|\hat{h}_2) = \bar{\lambda}_n(x|\hat{h}_2)+o_p(n^{-\frac{1}{2}})$$

with

$$\bar{\lambda}_n(x|\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}K\left(\frac{x-X_i}{\hat{h}_2}\lambda(X_i)^{\frac{1}{2}}\right)}{1-F(X_i)}.$$

Thus, (20) is equivalent to

$$\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \bar{\lambda}_n(x|\hat{h}_2)+T(x|\hat{h}_1,\hat{h}_2)+o_p\left(n^{-\frac{4}{9}}\right). \tag{21}$$

Now, (21) is implied by

$$\sup_{x\in[0,T]}\left|\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2)-\bar{\lambda}_n(x|\hat{h}_2)-T(x|\hat{h}_1,\hat{h}_2)\right| = o_p\left(n^{-\frac{4}{9}}\right). \tag{22}$$

Applying the definition of stochastic convergence we can rewrite (22) as

$$\lim_{n\to\infty}\mathrm{P}\left(\sup_{x\in[0,T]}\left|\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2)-\bar{\lambda}_n(x|\hat{h}_2)-T(x|\hat{h}_1,\hat{h}_2)\right|>\xi n^{-\frac{4}{9}}\right)=0,\quad \xi>0$$

To prove this, we define a relationship between the pilot estimate and the true hazard rate and we use this to write the adaptive estimate in terms of the ideal and some remainder terms. Then we work out the difference between these terms and T. First we split the adaptive into parts and then we investigate how T is related to those remaining terms. Lemmas used throughout the proof are proved in section A. Define $\delta(x)$ from the relation

$$\hat{\lambda}(x|\hat{h}_1)^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\{1+\delta(x)\}. \tag{23}$$

Also set $b(x|h) = \mu_2(x|h) - \lambda(x)$ and $\mathrm{D}(x|h) = \hat{\lambda}(x|h) - \mu_2(x|h)$. Then we can write (lemma A.1, pp. 28)

$$\hat{\lambda}(x|\hat{h}_1)^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\left\{1+\frac{\mathrm{D}(x|\hat{h}_1)+b(x|\hat{h}_1)}{\lambda(x)}\right\}^{\frac{1}{2}}. \tag{24}$$

Next we break the adaptive into parts by performing a Taylor expansion on the kernel, substituting back to the estimator and rearrange. Using (23) and Taylor series we write the kernel of the adaptive as

$$\begin{aligned}
K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\hat{\lambda}(u|\hat{h}_1)^{\frac{1}{2}}\right\} &= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}(1+\delta(u))\right\}\\
= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} &+ \left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\delta(u)K'\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\}\\
&\qquad + \delta_2(x,u)\\
= K\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} &+ \delta(u)L_1\left\{\left(\frac{x-u}{\hat{h}_2}\right)\lambda(u)^{\frac{1}{2}}\right\} + \delta_2(x,u) \qquad (25)
\end{aligned}$$

where

$$|\delta_2(x,u)| \le C_1\delta(u)^2\mathrm{I}\left(|x-X_i| \le C_2\hat{h}_2\right)$$

uniformly in $x, u \in (0,T)$. Here the indicator function $\mathrm{I}\left(|x-X_i| \le C_2\hat{h}_2\right)$ is introduced in order to exclude large values of u as this will ensure that observations away from the evaluation point will not have large effect on the estimate. Substituting this expression for the kernel to $\bar{\lambda}_n$, using (23) and rearranging gives,

$$\begin{aligned}
\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}&\frac{\lambda(X_i)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\
&+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}(\delta(X_i)+\delta^2(X_i))L_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\
&+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\delta(X_i)\lambda(X_i)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\
&+\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}+\delta(X_i)\lambda(X_i)^{\frac{1}{2}}}{1-F(X_i)}\delta_2(x,X_i).
\end{aligned}$$

Hence,

$$\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \bar{\lambda}_n(x|\hat{h}_2) + \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}\delta(X_i)L\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)} + \delta_3(x) \tag{26}$$

where

$$\begin{aligned}\delta_3(x) = \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}&\left\{\frac{\delta^2(X_i)L_1\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}+\right.\\ &\left.\frac{\delta(X_i)\lambda(X_i)^{\frac{1}{2}}K\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)} + \frac{\lambda(X_i)^{\frac{1}{2}}+\delta(X_i)\lambda(X_i)^{\frac{1}{2}}}{1-F(X_i)}\delta_2(x,X_i)\right\}\\ &\le C_1\left\{\sup_{X_i\in[0,T]}\delta^2(X_i)\right\}\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\mathrm{I}\left(|x-X_i|\le C_2\hat{h}_2\right)\end{aligned}$$

for a suitably chosen constant C_1 which includes the bound for the denominator, uniformly in $x\in(0,T)$. Since we already have $\bar{\lambda}_n$ in (26), we only need to form the T term from the last two terms of (26). Rewrite (24) as

$$\hat{\lambda}(x|\hat{h}_1)^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\left\{1+\delta_4(x)+\delta_5(x)\right\}$$

where

$$\delta_4(x) = \frac{\mathrm{D}(x|\hat{h}_1)+b(x|\hat{h}_1)}{2\lambda(x)},\quad \delta_5(x)\le C\left\{\mathrm{D}(x|\hat{h}_1)^2+b(x|\hat{h}_1)^2\right\}$$

uniformly in $x\in(0,T)$. Then, for $\delta=\delta_4+\delta_5$ (lemma A.2, pp. 28)

$$\bar{\lambda}_n(x|\hat{h}_1,\hat{h}_2) = \bar{\lambda}_n(x|\hat{h}_2) + \frac{1}{2}\varepsilon_1(x|\hat{h}_1,\hat{h}_2) + \frac{1}{2}\varepsilon_2(x|\hat{h}_1,\hat{h}_2) + \varepsilon_3(x|\hat{h}_1,\hat{h}_2) \tag{27}$$

where we define the ε_i, $i=1,2,3$ to be

$$\begin{aligned}\varepsilon_1(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{-\frac{1}{2}}\mathrm{D}(X_i|\hat{h}_1)L\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ \varepsilon_2(x|\hat{h}_1,\hat{h}_2) &= \frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{-\frac{1}{2}}b(X_i|\hat{h}_1)L\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}\\ |\varepsilon_3(x|\hat{h}_1,\hat{h}_2)| &\le C_1\varepsilon_4(x|\hat{h}_1,\hat{h}_2)\\ &\equiv \frac{C_1}{n\hat{h}_2}\left(\sup_{x\in[0,T]}\left\{\mathrm{D}(x|\hat{h}_1)^2+b(x|\hat{h}_1)^2\right\}\sum_{i=1}^{n}\mathrm{I}\left(|x-X_i|\le C_2\hat{h}_2\right)\right).\end{aligned}$$

Since we have proved so far that the adaptive estimator can be written in the form (27) we now need to work out the sum of the ε_i's, $i=1,2,3$. Specifically we want to show that

$$\frac{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_1(x|\hat{h}_1,\hat{h}_2)+\frac{1}{2}\varepsilon_2(x|\hat{h}_1,\hat{h}_2)+\varepsilon_3(x|\hat{h}_1,\hat{h}_2)-T(x|\hat{h}_1,\hat{h}_2)\right|}{n^{-\frac{4}{9}}}\xrightarrow{p}0. \tag{28}$$

For $\eta > \rho > 0$ let $\mathcal{H}_1 = \{h : \eta n^{-a-\frac{1}{5}} \leq h \leq \rho n^{a-\frac{1}{5}}\}$ and $\mathcal{H}_2 = \{h : \eta n^{-\frac{1}{9}} \leq h \leq \rho n^{\frac{1}{9}}\}$. Equivalently (lemma A.3, pp. 28), we show that

$$\sup_{x\in[0,T],h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2} \mathrm{P}\left\{|\varepsilon_i(x|h_1,h_2)| > \xi n^{-\frac{4}{9}}\right\} = O(n^{-r}), \quad r > 0 \tag{29}$$

for $i = 2, 4$ and

$$\sup_{x\in[0,T],h_1\in\mathcal{H}_1,h_2\in\mathcal{H}_2} \mathrm{P}\left\{|\varepsilon_1(x|h_1,h_2) - 2T(x|h_1,h_2)| > \xi n^{-\frac{4}{9}}\right\} = O(n^{-r}). \tag{30}$$

A crucial point of the proof will be the repeated use of inequality (21.5) of Burkholder [3]. In Burkholder's notation let $(f_1, f_2, \dots)$ be a martingale relative to $(\mathcal{F}_1, \mathcal{F}_2, \dots)$, a nondecreasing sequence of sub-σ-fields of $\mathcal{F}$. Let $(d_1, d_2, \dots)$ be the difference sequence of $f : f_n = \sum_{i=1}^n d_i$, $n \geq 1$. The square function and the maximal function of f are

$$S(f) = \left(\sum_{i=1}^{+\infty} \mathbb{E}\left\{d_i^2\right\}\right)^{\frac{1}{2}} \quad \text{and} \quad f^* = \sup_n |f_n|$$

respectively. Then, for a function Φ that is non-decreasing and continuous on $[0, +\infty)$ and satisfies $\Phi(2x) \leq c\Phi(x)$ the following inequality holds

$$\mathbb{E}\Phi(f^*) \leq c\mathbb{E}\Phi(S(f)) + C\sum_{i=1}^{+\infty}\mathbb{E}\Phi\left(|d_i|\right) \tag{31}$$

for some constant C. Starting with ε_4 we see that we need to bound functions D, b and the sum. We start with function D. Suppose that $\mathcal{J}$ is a finite set, subset of $(0, \mathrm{T})$ such that

$$\mathcal{J} = \{x \in (0,T) : \text{for some } y \in [0,T], |x-y| < \varepsilon\}, \varepsilon > 0.$$

An upper bound for $\mathrm{E}|\mathrm{D}(x|h_1)|^l$ for some appropriate constant C_1 will be (lemma A.4, pp. 29)

$$\mathbb{E}\left\{|\mathrm{D}(x|h_1)|^l\right\} \leq C_1\left(\frac{1}{nh_1}\right)^{\frac{l}{2}} \leq C_1\left(\frac{1}{n^{a+\frac{4}{5}}}\right)^{\frac{l}{2}}$$

uniformly in $x \in (0, T)$.

Suppose that the number of the elements of $\mathcal{J}$ increases at most algebraically fast in n. That is, we assume that the set $\mathcal{J}$ has $O(n^s)$ elements, where s is large but fixed. Working as in Stone [20] and using the fact that $\sum_i \mathbb{E}z_i \leq \{\#\mathcal{J}\}\sup\mathbb{E}z_i$ together with corollary 2.2 from Hall and Heyde ([6], pp. 19) we prove that the probability below is $O(n^{-r})$. Then we use Hölder continuity of the kernel to extend the result in the general case. We have

$$\mathrm{P}\left\{\sup_{x\in\mathcal{J}}\mathrm{D}(x|h_1)^2 > \xi n^{-\frac{4}{9}}\right\} \leq \{\#\mathcal{J}\}\left\{\xi^{-1}n^{\frac{4}{9}}\right\}^{\frac{1}{2}}\sup_{x\in\mathcal{J}}\mathbb{E}\left\{|\mathrm{D}(x|h_1)|^l\right\}.$$

Since the kernel is a Hölder continuous function we have that

$$|\mathrm{D}(x|h_1) - \mathrm{D}(y|h_1)| \leq c\,|x-y|^l.$$

From the definition of Hölder continuity we have that l can be any number greater than zero. This means that the distance between two successive elements of $\mathcal{J}$ can be as small or as big we like. Therefore, on choosing l large we get

$$\mathrm{P}\left\{\sup_{x\in(0,T)} \mathrm{D}(x|h_1)^2 > \xi n^{-\frac{4}{9}}\right\} = O\left(n^{-r}\right),\ r>0, \tag{32}$$

uniformly in $h_1 \in \mathcal{H}_1$. The sum can be bounded as in [7] by using exponential bounds

$$\begin{aligned}\mathrm{P}\left\{\sup_{x\in[0,T]} \frac{1}{h_2}\sum_{i=1}^{n} \mathrm{I}(|x-X_i|\le h_2) > C_4\right\}& \\ \le \mathrm{P}\left\{\sup_{x\in[0,T]} \frac{1}{h_2}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{|x-X_i|^2}{2h_2^2}} > C_3\right\} &= O(n^{-r})\end{aligned} \tag{33}$$

for $C_3 \le C_4$ finite positive constants. Also,

$$\begin{aligned}|b(x|h_1)| = |\mu_2(x|h_1) - \lambda(x)| = \left|\mathbb{E}\hat{\lambda}(x|h) - \lambda(x)\right| = \\ \left|\lambda(x) + \frac{h_1^2}{2!}\int z^2 K(z)\,dz + o(h^4) - \lambda(x)\right| \le C_5 h_1^2 \le \left(n^{a-\frac{1}{5}}\right)^2 = C_6 n^{2a-\frac{2}{5}}\end{aligned} \tag{34}$$

uniformly in $h_1 \in \mathcal{H}_1$ and $x \in (0,T)$. By (32), (33) and (34), (29) for $i=4$ is proved. We proceed with ε_2. We have,

$$\begin{aligned}&\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} = \\ &\frac{1}{\hat{h}_2}\mathbb{E}\int \frac{1}{n}\sum_{i=1}^{n} \frac{\lambda(z)^{-\frac{1}{2}} b(z|h_1)}{1-F(z)} L\left\{\left(\frac{x-z}{\hat{h}_2}\right)\lambda(z)^{\frac{1}{2}}\right\} f(z)\,dz \\ &= \int \lambda(x-h_2z)^{-\frac{1}{2}} b(x-h_2z|h_1) L\left\{z\lambda(x-h_2z)^{1/2}\right\}\lambda(x-h_2z)\,dz \\ &= \int \lambda(x-h_2z)^{\frac{1}{2}} b(x-h_2z|h_1) L\left\{z\lambda(x-h_2z)^{1/2}\right\}\,dz = O(h_1^2h_2^2)\end{aligned}$$

by lemma A.5, uniformly in $x \in (0,T)$. Thus, working as in the ε_4, assuming that l is sufficiently large and using lemma A.6,

$$\begin{aligned}\mathrm{P}\left\{|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| > \xi n^{-4/9}\right\} &\le \left(\xi^{-1}n^{4/9}\right)^l \mathbb{E}\left\{|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)|^l\right\} \\ &\le C(l)\xi^{-1} n^{\frac{4l}{9}}\left(\frac{1}{h_2}\right)^l\left\{(h_1^2h_2)^{\frac{l}{2}} + h_1^{2l}h_2^l\right\} = O(n^{-r}).\end{aligned}$$

This proves (29) for $i=2$. For ε_1 first define

$$\begin{aligned}m(u,v) &= \frac{\lambda(v)^{-\frac{1}{2}}\left\{\frac{K\left(\frac{v-u}{h_1}\right)}{1-F(v)} - h_1\mu_2(v|h_1)\right\} L\left\{\left(\frac{x-v}{h_2}\right)\lambda(v)^{\frac{1}{2}}\right\}}{1-F(v)}, \\ m_1(u) &= \mathbb{E}\left\{m(u,X_1)\right\} \text{ and } M(u,v) = m(u,v) - m_1(u).\end{aligned}$$

Then

$$\sum_{j=1}^{n} m_1(X_j) = \sum_{j=1}^{n} \mathbb{E}\left\{m(X_j, X_1)|X_j\right\}$$
$$= \sum_{j=1}^{n} t(X_j, x|h_1, h_2) = 2nh_1h_2T(x|h_1, h_2).$$

Therefore, (lemma A.7, pp. 33),

$$\varepsilon_1(x|h_1, h_2) = \frac{1}{n^2h_1h_2}\left\{\sum\sum_{i\neq j} M(X_i, X_j) + \sum_{i=1}^{n} M(X_i, X_i)\right\} + 2T(x|h_1, h_2) \tag{35}$$

Hence, it remains to show that for $\xi, r > 0$ we have

$$\sup_{x\in[0,T], h_1\in\mathcal{H}_1, h_2\in\mathcal{H}_2} \mathrm{P}\left\{\frac{1}{n^2h_1h_2}\left|\sum_{i=1}^{n} M(X_i, X_i)\right| > \xi n^{-\frac{4}{5}}\right\} = O(n^{-r}) \tag{36}$$

$$\sup_{x\in[0,T], h_1\in\mathcal{H}_1, h_2\in\mathcal{H}_2} \mathrm{P}\left\{\frac{1}{n^2h_1h_2}\left|\sum\sum_{i\neq j} M(X_i, X_j)\right| > \xi n^{-\frac{4}{5}}\right\} = O(n^{-r}) \tag{37}$$

We treat these two equations separately starting with (36).

$$\mathbb{E}\sum_{i=1}^{n} M^2(X_i, X_i) \le \mathbb{E}\sum_{i=1}^{n} m^2(X_i, X_i) \le$$
$$n\int \frac{\lambda(z)^{-1}}{(1-F(z))^4} K^2(0) L^2\left(\frac{x-z}{h_2}\lambda(z)^{\frac{1}{2}}\right) f(z)\, dz$$
$$= \int \frac{nK^2(0)}{(1-F(x))^3} L^2\left(\frac{x-z}{h_2}\lambda(z)^{\frac{1}{2}}\right) dz \le C_6 n h_2$$

with C_6 being a positive generic constant. Also,

$$\sum_{i=1}^{n} \mathbb{E}|M(X_i, X_i)|^l \le \sum_{i=1}^{n} \mathbb{E}|m(X_i, X_i)|^l = n\int \frac{\lambda(z)^{-\frac{l}{2}}}{(1-F(z))^{2l}} K^l(0) L^l\left(\frac{x-z}{h_2}\lambda(z)^{\frac{1}{2}}\right) f(z)\, dz$$

and thus,

$$\sum_{i=1}^{n} \mathbb{E}|M(X_i, X_i)|^l \le nC_7h_2$$

where C_7 is positive generic constant. By (31) with $\Phi(x) = x^l$,

$$f^* = \sup_{i=1,\dots,n} \sum_{i=1}^{n} M(X_i, X_i),\ d_i = M(X_i, X_i) - \mathbb{E}M(X_i, X_i)$$

and

$$S(f) = \left(\sum_{i=1}^{n} \mathbb{E}\left\{d_i\right\}^2\right)^{\frac{1}{2}}$$

gives

$$\mathbb{E}\left|\sum_{i=1}^{n} M(X_i, X_i)\right|^l \leq \mathbb{E}\left|\sum_{i=1}^{n}\left\{\sup_{i=1,\ldots,n}\sum_{i=1}^{n} M(X_i, X_i)\right\}\right|^l$$

$$\leq \left|\sum_{i=1}^{n}\mathbb{E}\left\{M(X_i, X_i) - \mathbb{E}M(X_i, X_i)\right\}^2\right|^{\frac{l}{2}} + \sum_{i=1}^{n}\left|\mathbb{E}\left\{M(X_i, X_i) - \mathbb{E}M(X_i, X_i)\right\}\right|^l$$

$$\leq \left|\sum_{i=1}^{n}\mathbb{E}\left\{M(X_i, X_i)\right\}^2\right|^{\frac{l}{2}} + \sum_{i=1}^{n}\left|\mathbb{E}\left\{M(X_i, X_i)\right\}\right|^l$$

$$\leq (n^2h_1h_2)^{-l}((C_6nh_2)^{\frac{l}{2}} + C_7nh_2).$$

Hence,

$$\mathrm{P}\left\{\frac{1}{n^2h_1h_2}\left|\sum_{i=1}^{n} M(X_i, X_i)\right| > \xi n^{-\frac{4}{5}}\right\} \leq C_7(l)\xi^{-l}\left(\frac{1}{n^2h_1h_2}\right)^l$$

$$\times\left\{\mathbb{E}\left|\sum_{i=1}^{n}\left\{M(X_i, X_i) - \mathbb{E}M(X_i, X_i)\right\}^2\right|^{\frac{l}{2}} + \sum_{i=1}^{n}\mathbb{E}\left|M(X_i, X_i) - \mathbb{E}M(X_i, X_i)\right|^l\right\}$$

$$\leq C_8(l)\xi^{-l}\left(\frac{1}{n^2h_1h_2}\right)^l\left\{(nh_2)^{l/2} + nh_2\right\}.$$

Choosing l sufficiently large completes the proof of (36). For the proof of (37) first denote with $N(x, y)$ any of the two functions $M(x, y)$, $M(y, x)$. Since the $N(X_i, X_j)$ are identically distributed and since

$$\sum\sum_{i<j} M(X_i, X_j) + \sum\sum_{j<i} M(X_i, X_j) = \sum\sum_{i\neq j} M(X_i, X_j)$$

we can examine only the case where $\sum\sum_{i\neq j} M(X_i, X_j)$ is replaced by $\sum\sum_{i\leq j} N(X_i, X_j)$. Set

$$Z_j = \sum_{1\leq i\leq j-1} N(X_i, X_j)$$

and note that

$$\mathbb{E}\{N(X_i, X_j)|X_i\} = \mathbb{E}\left\{m(X_i, X_j) - m_1(X_i)|X_i\right\} = 0$$

for $i \neq j$ so the Z_j's are martingale differences. Also,

$$\sum_{j=1}^{n}\mathbb{E}\{Z_j\}^2 = \sum_{j=1}^{n}\mathbb{E}\left\{\sum_{i=1}^{j-1} N(X_i, X_j)\right\}^2 = \frac{n(n-1)}{2}\mathbb{E}\left\{N^2(X_1, X_2)\right\}$$

and therefore

$$\mathbb{E}N^2(X_1, X_2) \leq \iint \frac{f(u)}{(1-F(v))^3}K^2\left(\frac{u-v}{h_1}\right)L^2\left(\frac{x-v}{h_2}\lambda(v)^{\frac{1}{2}}\right)du\,dv \leq C_{10}h_1h_2.$$

Notice that Z_j conditional on X_j is a sum of independent and identically distributed random variables. Thus we can use Rosenthal's inequality to obtain a bound for $\mathbb{E}|Z_j|^l$:

$$\mathbb{E}\left\{|Z_j|^l\right\} \leq$$

$$C_9\mathbb{E}\left(\left\{(j-1)\mathbb{E}\left\{\left|N^2(X_{j-1},X_j)\right|\Big|X_j\right\}\right\}^{\frac{l}{2}} + (j-1)\mathbb{E}\left\{|N(X_{j-1},X_j)|^l\Big|X_j\right\}\right)$$

and since we have

$$\mathbb{E}\left\{\left|N^2(X_{j-1},X_j)\right|\Big|X_j\right\} \leq$$

$$\int \frac{1}{(1-F(v))^3}K^2\left(\frac{u-v}{h_1}\right)L^2\left(\frac{x-v}{h_2}\lambda(v)^{\frac{1}{2}}\right)dv \leq C_{10}h_2$$

and

$$\mathbb{E}\left\{|N(X_{j-1},X_j)|^l\Big|X_j\right\} \leq$$

$$\int \frac{\lambda^{-\frac{l}{2}}(v)}{(1-F(v))^{2l}}K^l\left(\frac{u-v}{h_1}\right)L^l\left(\frac{x-v}{h_2}\lambda^{\frac{1}{2}}(v)\right)f(v)\,dv \leq C_{11}h_2$$

we get that

$$\mathbb{E}\left\{|Z_j|^l\right\} \leq C_{12}\left((jh_2)^{\frac{l}{2}} + jh_2\right).$$

Finally from (31) with $\Phi(x) = x^l$,

$$f^* = \sup_{i=1,\ldots,n}\sum\sum_{1\leq i<j\leq n} N(X_i,X_j),\ d_i = Z_j \text{ and } S(f) = \left(\sum_{j=1}^{n}\mathbb{E}\left\{Z_j\right\}^2\right)^{\frac{1}{2}}$$

we have that

$$\mathbb{E}\left|\sum\sum_{1\leq i<j\leq n} N(X_i,X_j)\right|^l \leq C_{13}\left((n^2h_1h_2)^{\frac{l}{2}} + n(nh_2)^{\frac{l}{2}}\right)$$

which completes the proof on choosing l large. □

Now, define

$$S(x|h_2) = \tilde{\lambda}_n(x|h_2) - \mathbb{E}\tilde{\lambda}_n(x|h_2).$$

The next theorem is used to evaluate how much we lose in practice by using the adaptive estimator instead of the ideal. It shows that the remainder term, $T(x|h_1,h_2)$, is of the same order as the difference between $\tilde{\lambda}_n(x|h)$ and the true hazard $\lambda(x)$.

Theorem 3.2. *Assume that the kernel satisfies conditions A1, A2 and that it has two bounded derivatives. Let $\lambda > 0$ be bounded on $(0,T)$ and continuous at $x \in (0,T)$. Suppose that for non-random bandwidths h_1, h_2 satisfying $n^\varepsilon \max(h_1,h_2) \to 0$, $n^{1-\varepsilon}\min(h_1,h_2) \to \infty$ for some $\varepsilon > 0$ and $h_1h_2^{-1} \to 0$ we have $\hat{h}_1/h_1 \to 1$ and $\hat{h}_2/h_2 \to 1$ in probability. Then*

$$\sqrt{nh_2}\left(S(x|\hat{h}_2), T(x|\hat{h}_1,\hat{h}_2)\right) \xrightarrow{d} (N_1,N_2)$$

where (N_1, N_2) *is a bivariate normal distribution with mean 0 and covariance*

$$\mathbb{V}\mathrm{ar}(N_1) = \frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\int K^2, \mathbb{V}\mathrm{ar}(N_2) = \frac{1}{4}\frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\int L^2,$$
$$\mathbb{C}\mathrm{ov}(N_1, N_2) = \frac{1}{2}\frac{\lambda(x)^{\frac{3}{2}}}{1-F(x)}\int KL.$$

Proof. To avoid long technical details we consider here nonrandom bandwidths to prove asymptotic normality and then to establish asymptotic variances and covariances. The case of random bandwidths can be treated by expanding the arguments in Abramson (1982) to the hazard case.

The proof of asymptotic normality is based on the Hajek projection method, [9], which essentially extents the scope of the central limit theorem to sums that are asymptotically equivalent to sums of independent random variables. Here, we are concerned with the statistic S. Applying Hajek's idea we see that the result will follow if we approximate S by its projection, say $\hat{S}$, on the subspace of all such sums of independent terms. Recall the definition of S,

$$S = \tilde{\lambda}_n(x|h) - \mathbb{E}\tilde{\lambda}_n(x|h).$$

It is immediately seen that it is equivalent to prove that the standardized version of $\tilde{\lambda}_n$ has asymptotically a standard normal distribution. To prove that we follow the proof of Tanner and Wong, [18]. Write

$$S = \sum_{i=1}^{n} V_i, \quad V_i = \frac{1}{nh}\frac{\lambda(X_i)^{\frac{1}{2}} K\left(\frac{x-X_i}{h}\lambda(X_i)^{\frac{1}{2}}\right)}{1-F_n(X_i)}$$

and set

$$\hat{S} = \sum_{i=1}^{n} \mathbb{E}(S|X_i) - (n-1)\mathbb{E}S.$$

Then we easily see that $\mathbb{E}\hat{S} = \mathbb{E}S$ and $\mathbb{E}(S-\hat{S})^2 = \mathbb{V}\mathrm{ar}(S) - \mathbb{V}\mathrm{ar}(\hat{S})$. Now,

$$\mathbb{E}(V_i|X_i) = \frac{1}{n}Z(X_i) \text{ and for } i \neq j,$$
$$\mathbb{E}(V_j|X_i) = \frac{1}{n-1}\frac{1}{h}\int \left(1-F^{n-1}(y)\right) K\left(\frac{x-y}{h}\lambda(y)^{\frac{1}{2}}\right)\lambda(y)^{\frac{3}{2}}\,dy$$
$$+\frac{1}{n(n-1)}Q(X_i)$$

where

$$Z(X_i) = \frac{1-F^n(X_i)}{1-F(X_i)}K\left(\frac{x-X_i}{h}\lambda(X_i)^{\frac{1}{2}}\right)\lambda(X_i)^{\frac{1}{2}}$$
$$Q(X_i) = -\int \frac{1-F^n(y)-nF^{n-1}(y)(1-F(y))}{1-F(y)}K\left(\frac{x-y}{h}\lambda(y)^{\frac{1}{2}}\right)\lambda(y)^{\frac{3}{2}} I_{\{y\le X_i\}}\,dy.$$

Now,

$$\begin{aligned}\hat{S} - \mathbb{E}\hat{S} &= \sum_{i=1}^{n} \{\mathbb{E}(V_i|X_i) - (n-1)\mathbb{E}(V_j|X_i) - \mathbb{E}S\} \\ &= \sum_{i=1}^{n} \left\{ \frac{1}{n} Z(X_i) + \frac{1}{n} Q(X_i) + R_n \right\}\end{aligned}$$

where

$$R_n = -\int F^{n-1}(y) K\left(\frac{x-y}{h} \lambda(y)^{\frac{1}{2}} \right) \lambda(y)^{\frac{1}{2}} f(y)\, dy.$$

It can be easily shown that

$$|Q| = O(\log n), |R_n| = O\left(\frac{1}{n(n+1)}\right) \text{ and } \mathbb{E}|Z(X_i)|^r = \frac{\lambda(x)^{\frac{3}{2}}}{h^{r-1}} \int K^r/(1-F)^{r-1}.$$

Utilizing these results we can show that $\hat{S}$ and S have the same asymptotic distribution. For that consider

$$\mathbb{V}\text{ar}(\hat{S}) = n\mathbb{V}\text{ar}\left(\frac{1}{n} Z + \frac{1}{n} Q + R_n \right) = \frac{1}{n} \frac{\lambda(x)^{\frac{3}{2}}}{h} \int K^2 + o\left(\frac{1}{nh}\right)$$

therefore, in view of the variance of S we get that $\mathbb{V}\text{ar}(\hat{S})/\mathbb{V}\text{ar}(S) \to 1$. Hence,

$$\mathbb{E}\left(\frac{\hat{S} - \mathbb{E}\hat{S}}{\sqrt{\mathbb{V}\text{ar}(\hat{S})}} - \frac{S - \mathbb{E}S}{\sqrt{\mathbb{V}\text{ar}(\hat{S})}} \right)^2 = \frac{\mathbb{E}\left(\hat{S} - S\right)^2}{\mathbb{V}\text{ar}(\hat{S})} = \frac{\mathbb{V}\text{ar}(\hat{S}) - \mathbb{V}\text{ar}(S)}{\mathbb{V}\text{ar}(\hat{S})} \to 0.$$

Finally, in order to show that the asymptotic distribution of the standardized statistic $\hat{S}$ is standard normal we use Lyapunov's theorem according to which, since R_n is negligible, a sufficient condition is that

$$\frac{n\mathbb{E}\left|\frac{1}{n}Z + \frac{1}{n}Q\right|^3}{\sqrt{\mathbb{V}\text{ar}(\hat{S})^3}} \to 0$$

which is already established because using the bounds for Q and Z the above quantity is $O((nh)^{-\frac{1}{2}})$.

We now proceed to establish asymptotic variances and covariances. Define

$$\begin{aligned} m(u,v) &= \frac{\lambda(v)^{-\frac{1}{2}} \left\{ \frac{K\left(\frac{v-u}{h_1}\right)}{1-F(v)} - h_1 \mu(v|h_1) \right\} L\left\{ \left(\frac{x-v}{h_2}\right) \lambda(v)^{\frac{1}{2}} \right\}}{1-F(v)}, \\ m_1(u) &= \mathbb{E}\{m(u, X_j)\} \end{aligned}$$

for any $j = 1, 2, \ldots, n$. Then we can write

$$m_1(u) = m_2(u) - m_3$$

where

$$m_2(u) = \int \frac{\lambda^{1/2}(v)}{1-F(v)} K\left(\frac{u-v}{h_1}\right) L\left(\frac{x-v}{h_2}\lambda^{1/2}(v)\right) dv$$
$$= h_2 \int \frac{\lambda^{1/2}(x-h_2 z)}{1-F(x-h_2 z)} K\left(\frac{u-x+h_2 z}{h_1}\right) L\left(z\lambda^{1/2}(x-h_2 z)\right) dz$$

and

$$m_3 = h_1 h_2 \int \mu(x - vh_2|h_1)\lambda^{1/2}(x-vh_2) L\left(v\lambda^{1/2}(x-vh_2)\right) dv = O(h_1 h_2).$$

Thus, $T(x|h_1,h_2) = \frac{1}{2nh_1h_2}\sum_{i=1}^n m_2(X_i) - \frac{1}{2h_1h_2} m_3$ and its asymptotic variance will now equal to $\frac{1}{4}\frac{1}{n^2h_1^2h_2^2}\sum_{i=1}^n \mathbb{V}\mathrm{ar}\left\{m_2(X_i)\right\}$. Let $x - v = h_2 z$. Then,

$$m_2(u) = h_2 \int \frac{\lambda^{\frac{1}{2}}(x-h_2 z)}{1-F(x-h_2 z)} K\left(\frac{x-h_2 z-u}{h_1}\right) L\left(z\lambda^{\frac{1}{2}}(x-h_2 z)\right) dz \tag{38}$$

and so

$$\mathbb{E}m_2(X) = h_2 \times \int\left(\int \frac{\lambda^{\frac{1}{2}}(x-h_2 z)}{1-F(x-h_2 z)} K\left(\frac{x-h_2 z-w}{h_1}\right) L\left(z\lambda^{\frac{1}{2}}(x-h_2 z)\right) dz\right) f(w)\, dw$$

with the change of variable $x - w = h_1 r$ the mean becomes

$$\mathbb{E}m_2(X) = h_2 h_1 \times \int\left(\int \frac{\lambda^{\frac{1}{2}}(x-h_2 z)}{1-F(x-h_2 z)} K\left(r-h_2 z\right) L\left(z\lambda^{\frac{1}{2}}(x-h_2 z)\right) dz\right) f(x-h_1 r)\, dr.$$

Again, using change of variables similar to (3)–(6) and working as in lemma A.5, the above expression reduces to

$$\mathbb{E}m_2(X) = h_2 h_1 \lambda(x) \iint K(u)\, L(w)\, du\, dw = O(h_1 h_2), \tag{39}$$

which is negligible. Now rewrite $m_2(u)$ by substituting $z = v_1 h$ where $h = h_1 h_2^{-1}$ as

$$m_2(u) = h_1 \int \frac{\lambda^{1/2}(x-h_1 v_1)}{1-F(x-h_1 v_1)} K\left(\frac{u-x+h_1 v_1}{h_1}\right) L\left(hv_1\lambda^{1/2}(x-h_1 v_1)\right) dv_1.$$

Then

$$\mathbb{E}\left\{m_2^2(X)\right\} = h_1^2 \iiint \frac{\lambda^{1/2}(x-h_1 v_1)}{1-F(x-h_1 v_1)} K\left(\frac{u-x+h_1 v_1}{h_1}\right)$$
$$\times L\left(hv_1\lambda^{1/2}(x-h_1 v_1)\right) \frac{\lambda^{1/2}(x-h_1 v_2)}{1-F(x-h_1 v_2)} K\left(\frac{u-x+h_1 v_2}{h_1}\right)$$
$$\times L\left(hv_2\lambda^{1/2}(x-h_1 v_2)\right) f(u)\, dv_1\, dv_2\, du.$$

By substituting $h_1 w = u - x + h_1 v_1$ we get

$$\begin{aligned}\mathbb{E}\left\{m_2^2(X)\right\} = \\ h_1^3 \iiint & \frac{\lambda^{1/2}(x-h_1v_1)}{1-F(x-h_1v_1)} K(w) L\left(hv_1\lambda^{1/2}(x-h_1v_1)\right) \frac{\lambda^{1/2}(x-h_1v_2)}{1-F(x-h_1v_2)} \\ & \times K(w-v_1+v_2) L\left(hv_2\lambda^{1/2}(x-h_1v_2)\right) f(x+h_1w-h_1v_1)\, dv_1\, dv_2\, dw.\end{aligned}$$

Now set $Q(x) = \frac{\lambda^{1/2}(x)}{1-F(x)}$. Then expand Q and f in Taylor series around x, and put $a = \lambda^{1/2}(x)$ to get

$$\begin{aligned}\mathbb{E}\left\{m_2^2(X)\right\} &\simeq \\ h_1^3 Q^2(x) f(x) &\iiint K(w) L\left\{ahv_1)\right\} K(w-v_1+v_2) L(ahv_2)\, dv_1\, dv_2\, dw \\ = h_1^3 Q^2(x) f(x) &\iiint K(w) L\left\{ah(v_2+z))\right\} K(w+z) L(ahv_2)\, dz\, dv_2\, dw.\end{aligned}$$

Next, set $ahv_2 = u$ and after some algebra,

$$\mathbb{E}m_2^2(X) \simeq h_1^2 h_2 \frac{\lambda^{3/2}(x)}{1-F(x)} \int L^2(u)\, du, \tag{40}$$

By (39) and (40) the asymptotic variance of $T(x|h_1,h_2)$ follows.

Since $\mathbb{E}\left\{S(x|h_2)\right\} = 0$, the covariance will be calculated from

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(S(x|h_2), T(x|h_1,h_2)\right) &= \mathbb{E}\left\{S(x|h_2)T(x|h_1,h_2)\right\} \\ &= \mathbb{E}\left\{\tilde{\lambda}_n(x|h_2)T(x|h_1,h_2)\right\} - \mathbb{E}\tilde{\lambda}_n(x|h_2)\mathbb{E}T(x|h_1,h_2).\end{aligned} \tag{41}$$

For the second term of the right hand side of the last equation, considering only the leading term of the product of the means of $\tilde{\lambda}_n$ and m_2 (respectively from Theorem 2.1 and equation (39)), we have

$$\mathbb{E}\tilde{\lambda}_n(x|h_2)\mathbb{E}m_2(X_i) \simeq \lambda^2(x) h_1 h_2 \iint K(u)L(w)dudw,$$

and thus

$$\mathbb{E}\tilde{\lambda}_n(x|h_2)\mathbb{E}T(x|h_1,h_2) \simeq \frac{\lambda^2(x)}{2} \iint K(u)L(w)dudw. \tag{42}$$

To compute $\mathbb{E}\left\{\tilde{\lambda}_n(x|h_2)T(x|h_1,h_2)\right\}$, separate the diagonal and non diagonal terms to write,

$$\begin{aligned}\mathbb{E}\left\{\tilde{\lambda}_n(x|h_2)T(x|h_1,h_2)\right\} &= \\ & \frac{1}{2nh_1h_2}\frac{1}{nh_2}\mathbb{E}\left\{\sum_{i=1}^n m_2(X_i)\lambda(X_i)^{1/2}\frac{K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\} \\ & + \frac{1}{2nh_1h_2}\frac{1}{nh_2}\mathbb{E}\left\{\sum\sum_{i\neq j} m_2(X_j)\lambda(X_i)^{1/2}\frac{K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\}.\end{aligned} \tag{43}$$

For the diagonal term on the right hand side of (11) observe that

$$\frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n} m_2(X_i)\lambda(X_i)^{1/2}\frac{K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\}$$
$$= \mathbb{E}\left\{\sum_{i=1}^{n} m_2(X_{(i)})\frac{\lambda^{1/2}(X_{(i)})K\left(\frac{x-X_{(i)}}{h_2}\lambda^{1/2}(X_{(i)})\right)}{n-i+1}\right\}.$$

Using the change of variables $x - v = h_2 z$

$$\mathbb{E}\left\{\sum_{i=1}^{n} m_2(X_{(i)})\frac{\lambda^{1/2}(X_{(i)})K\left(\frac{x-X_{(i)}}{h_2}\lambda^{1/2}(X_{(i)})\right)}{n-i+1}\right\} =$$
$$\sum_{i=1}^{n}\int m_2(u)\frac{\lambda^{1/2}(u)K\left(\frac{x-u}{h_2}\lambda^{1/2}(u)\right)}{n-i+1}\binom{n}{i-1}F^{i-1}(u)(1-F(u))^{n-i}f(u)\,du \simeq$$
$$\int m_2(u)\lambda^{1/2}(u)K\left(\frac{x-u}{h_2}\lambda^{1/2}(u)\right)\frac{1-F^n(u)}{1-F(u)}f(u)\,du =$$
$$h_2\iint \frac{\lambda^{1/2}(x-h_2z)}{1-F(x-h_2z)}K\left(\frac{x-h_2z-u}{h_1}\right)L\left(z\lambda^{1/2}(x-h_2z)\right)$$
$$\times\, \lambda^{1/2}(u)K\left(\frac{x-u}{h_2}\lambda^{1/2}(u)\right)\lambda(u)(1-F^n(u))\,dz\,du.$$

Set $hh_2 = h_1$, $v_1 h = z$ and $x - u = h_1 r$ and note that $F^n(u) = o(n^{-1})$. This gives,

$$\frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n} m_2(X_i)\frac{\lambda^{1/2}(X_i)K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\} \simeq$$
$$h_1^2\iint \frac{\lambda^{1/2}(x-h_1v_1)}{1-F(x-h_1v_1)}K(r-v_1)L\left(hv_1\lambda^{1/2}(x-hv_1)\right)$$
$$\times\, \lambda^{\frac{3}{2}}(x-h_1r)K\left(hr\lambda^{1/2}(x-h_1r)\right)\,dr\,dv_1.$$

Expanding $Q(=\frac{\lambda^{1/2}}{1-F})$ and $\lambda^{3/2}$ in Taylor series around x and setting $a = \lambda^{1/2}(x)$,

$$\frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n} m_2(X_i)\frac{\lambda^{1/2}(X_i)K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\}$$
$$\simeq h_1^2\frac{\lambda^2(x)}{1-F(x)}\iint K(r-v_1)L\left(hv_1a\right)K\left(hra\right)\,dr\,dv_1.$$

Set $hv_1a = u$. Then,

$$
\begin{aligned}
(44) \quad \frac{1}{n}\mathbb{E}\left\{\sum_{i=1}^{n} m_2(X_i)\frac{\lambda^{1/2}(X_i)K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\} \\
\simeq h_1h_2\frac{\lambda^{3/2}(x)}{1-F(x)}\frac{1}{ha}\iint K\left(\frac{hra-u}{ha}\right)L\left(u\right)K\left(hra\right)\,dr\,du \\
= h_1h_2^2\frac{\lambda^{3/2}(x)}{1-F(x)}\int K(u)L(u)\,du.
\end{aligned}
$$

For the non diagonal term in (43), with similar arguments and algebra, but using the joint density of $(X_{(i)}, X_{(j)})$, one can derive that

$$
\begin{aligned}
(45) \quad \frac{1}{n}\mathbb{E}\left\{\sum\sum_{i\neq j} m_2(X_j)\lambda^{1/2}(X_i)\frac{K\left(\frac{x-X_i}{h_2}\lambda^{1/2}(X_i)\right)}{1-F_n(X_i)}\right\} = \\
h_1h_2^2\frac{\lambda^2(x)}{2}\iint K(u)L(w)\,du\,dw.
\end{aligned}
$$

Now, combining (41)-(45) we get

$$
\mathbb{C}\mathrm{ov}\left(S(x|h_2), T(x|h_1,h_2)\right) \simeq \frac{1}{2}\frac{\lambda^{\frac{3}{2}}(x)}{1-F(x)}\int K(u)\mathrm{L}(u)\,du.
$$

□

An immediate conclusion that can be drawn is that the remainder term does not change the rate of convergence, but it does prevent the adaptive estimator from achieving the same first-order properties as the ideal as explained in Remark 3.1 below.

Remark 3.1. *Since the limiting pair (N_1, N_2) has zero mean, we conclude that $\hat{\lambda}_n(x|h_1,h_2)$ and $\tilde{\lambda}_n(x|h)$ have the same asymptotic bias. Further, using the properties of kernels K and L it is not difficult to see that the covariance between S and T is positive. Hence, we conclude that the adaptive estimator has larger asymptotic variance than the ideal.*

Remark 3.2. *A consequence of Theorem 3.2 is that dependence of $T(x|\hat{h}_1,\hat{h}_2)$ on $\hat{h}_1$ vanishes in the limit. Therefore, the use of a more accurate estimator than $\hat{\lambda}(x|h)$ as pilot will not affect the asymptotic properties of $\hat{\lambda}_n(x|h_1,h_2)$.*

Remark 3.3. *The condition $h_1h_2^{-1} \to 0$ of Theorem 3.2 provides a guidance to choose the bandwidths. Practically we would choose a bigger value for h_2 than for h_1, i.e. $h_1 \sim n^{-\frac{1}{5}}$ and $h_2 \sim n^{-\frac{1}{9}}$ because in this way we achieve the fastest rate of convergence to the true hazard rate. For detailed discussion of this issue in the density estimation setting see Hall and Marron (1988).*

4. Numerical examples.

In this section we use distributional as well as real life data to exhibit the practical performance of the variable bandwidth estimator. To implement a kernel based estimator we need to specify the form of the kernel function and a bandwidth selection method. Of the two, most important is bandwidth selection as bandwidth determines the asymptotic properties of the estimate. The method we use here involves plotting the estimate for different values of h and choosing the curve which is 'most pleasing to the eye'. This bandwidth selection method is typically referred to as 'subjective choice' and it has the advantage that it 'may well give more insight into the data than merely a single automatically produced curve' (Silverman (1986), pp. 44).

We have developed an interactive program, essentially a Graphical User Interface (GUI), to implement the subjective bandwidth choice method for estimator $\hat{\lambda}_n(x|h_1, h_2)$ and the estimates which will be used for comparison. The code of this tool together with instructions and implementation examples can be found on statlib under the Get software / XLispStat Archive section, in the file "VBHRE.zip".

Using the Graphical User Interface for bandwidth choice we proceed now to illustrate estimator $\hat{\lambda}_n(x|h_1, h_2)$ in action. As in the case of real life data we don't know the true curve, we use for comparison the estimator

$$\hat{\lambda}_a(x|h_1, h_2) = \frac{\tilde{f}(x|h_1, h_2)}{1 - \hat{F}(x)} \tag{46}$$

where

$$\tilde{f}(x|h_1, h_2) = n^{-1} h_2^{-1} \sum_{i=1}^{n} \hat{f}(X_i|h_1)^{\frac{1}{2}} K \left\{ h_2^{-1} \hat{f}(X_i|h_1)^{\frac{1}{2}} (x - X_i) \right\},$$

$$\hat{f}(x|h_1) = n^{-1} h_1^{-1} \sum_{i=1}^{n} K \left(\frac{x - X_i}{h_1} \right),$$

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \mathcal{K} \left(\frac{x - X_i}{h} \right)$$

and $\mathcal{K}(u)$ is the cumulative distribution function of the kernel,

$$\mathcal{K}(u) = \int_{-\infty}^{u} K(v)\, dv.$$

This estimate has been discussed in Silverman (1986).

In Fig. 1 we estimate the hazard rate from a sample of 800 observations from the standard exponential distribution. The line parallel to the horizontal axis is the true hazard function, the solid line is estimator $\hat{\lambda}_n(x|h_1, h_2)$ and the dashed line is estimator $\hat{\lambda}(x|h_1)$. Both estimates use the Epanechnikov kernel. For both estimates we used equation (3.31) in Silverman (1986) to determine the bandwidth of the pilot estimators (bandwidth h_1 in (1) and (18)). To select bandwidth h_2 (bandwidth for the kernels in the adaptive estimates) we employed the Graphical User Interface and chose the bandwidth which results in a curve which is 'most pleasing to the eye'. This results in a bandwidth $h_2 = 0.17$ for estimator $\hat{\lambda}_n(x|h_1, h_2)$. It is apparent that the adaptive estimate has reduced bias compared to the

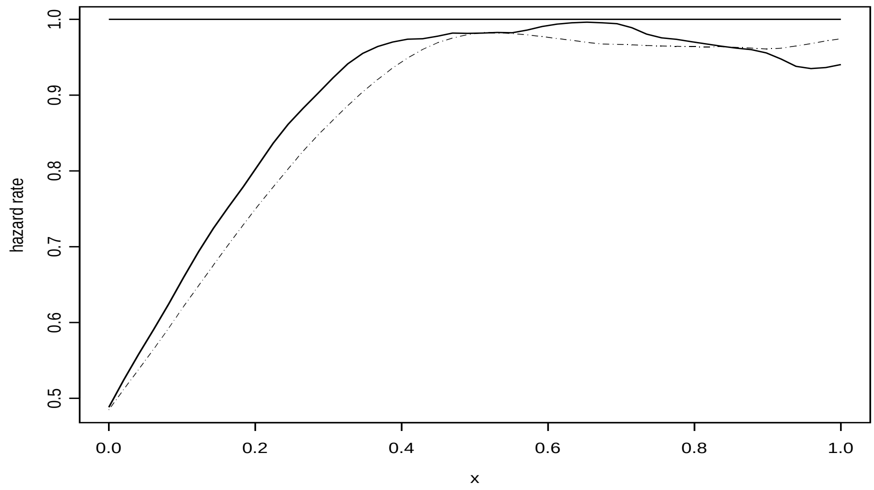

FIGURE 1. Comparison of $\hat{\lambda}_n(x|h_1, h_2)$ (solid line) with estimator $\hat{\lambda}(x|h_1)$ (dashed line). The parallel line to the horizontal axis is the real curve.

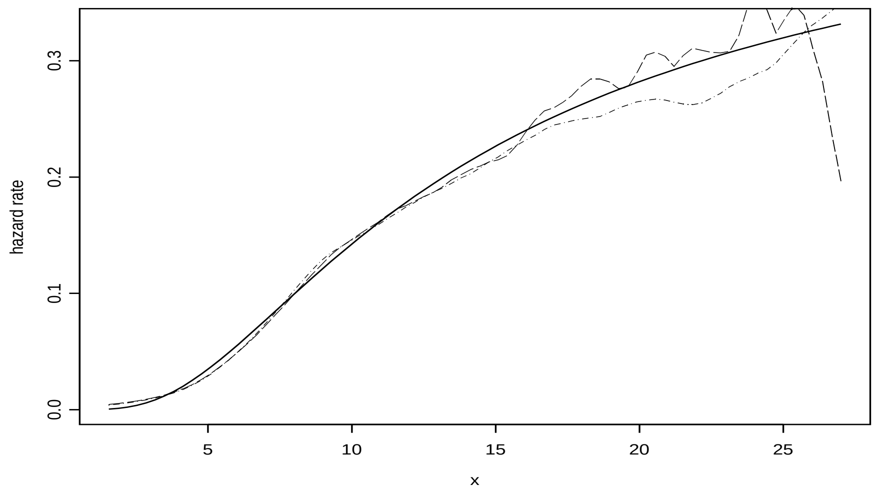

FIGURE 2. Comparison of $\hat{\lambda}_n(x|h_1, h_2)$ (dashed line) with estimator $\hat{\lambda}_a(x|h_1, h_2)$ (dotted - dashed line). The solid line is the real curve.

bias of the standard second order kernel estimate. As we see from the plot the adaptive estimate performs better than estimator $\hat{\lambda}(x|h_1)$ from the start and up to about $x = 0.9$.

Next, in Fig. 2 we compare estimators $\hat{\lambda}_n(x|h_1, h_2)$ and $\hat{\lambda}_a(x|h_1, h_2)$ by estimating the hazard rate from a sample of 1000 observations from the χ^2_{12} distribution. Again we use the Epanechnikov kernel for both estimates and we choose the bandwidth for the pilot estimators by the Silverman default rule. The bandwidths for the kernels of the adaptive estimates are

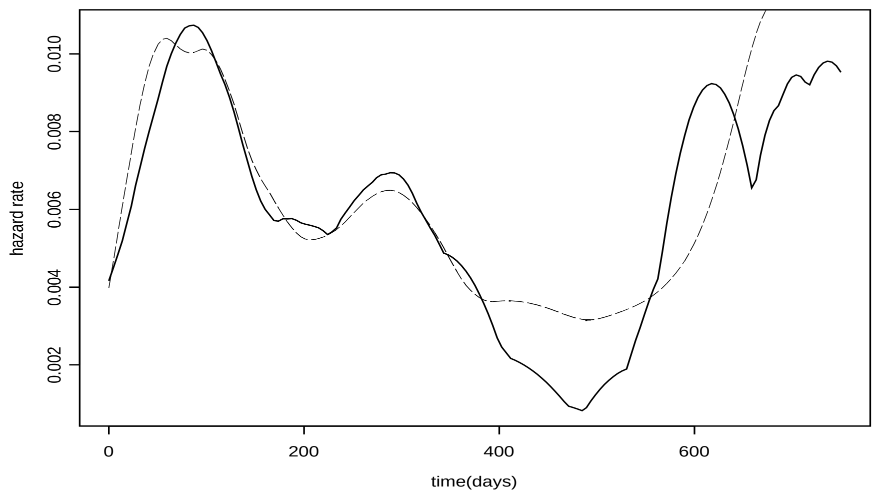

FIGURE 3. Comparison of $\hat{\lambda}_n(x|h_1, h_2)$ (solid line) with estimator $\hat{\lambda}_a(x|h_1, h_2)$ (dashed line).

chosen by the GUI. We see that the estimators perform similarly from the start until about $x = 16$ and from that point on estimator $\hat{\lambda}_n(x|h_1, h_2)$ seems to be nearer to the true curve.

In Fig. 3 we use the suicide data, Silverman (1986), to see the performance of the estimates on a real life application. The data are lengths of treatment spells (in days) of control patients in a suicide study. The solid line is estimator $\hat{\lambda}_n(x|h_1, h_2)$ and the dashed line is $\hat{\lambda}_a(x|h_1, h_2)$. Bandwidths for the pilot estimates were chosen by using equation (3.31) in Silverman (1986) and bandwidths for the kernels of the adaptive estimates were chosen with the subjective choice method. The method gave bandwidth $h_2 = 2.3$ for $\hat{\lambda}_n(x|h_1, h_2)$ and $h_2 = 4.5$ for $\hat{\lambda}_a(x|h_1, h_2)$. In both cases we used the biweight kernel. We see that in general both estimates perform similarly. Both curves suggest a fall in hazard in approximately $x = 100$ to $x = 500$ and a subsequent increase from $x = 500$ onwards. We have to note here the unreliability of such estimates for large values of x because of the denominator in their variance formulae. That is, the increase for $x > 500$ could be just random effect. However the two estimates have some differences such as at about $x = 100$ where $\hat{\lambda}_a$ suggests a bimodal structure in contrast with $\hat{\lambda}_n$ which shows a unimodal curve. Moreover at about $x = 500$, $\hat{\lambda}_n$ suggests a bigger drop in hazard than $\hat{\lambda}_a$.

5. CONCLUSIONS.

We showed that in the case of kernel-based nonparametric hazard rate estimation, a smoothing parameter inversely proportional to the square root of the true hazard rate leads to a mean square error rate of order $n^{-8/9}$. We discussed the practical implementation of such an estimate and applied it to distributional and real life data.

However the results here leave open two important issues. First, an extension of this work in the case of randomly right-censored data would be particularly useful as this is frequently the case in lifetime studies. Second, an automatic bandwidth selection rule for the adaptive estimate would be particularly interesting since this would make the method useful to practitioners.

Appendix A. Lemmas.

Lemma A.1 (Equation (24)).

$$\lambda(x)^{\frac{1}{2}}\left[1+\frac{\mathrm{D}(x|\hat{h}_1)+b(x|\hat{h}_1)}{\lambda(x)}\right]^{\frac{1}{2}} =$$

$$\lambda(x)^{\frac{1}{2}}\left[1+\frac{\hat{\lambda}(x|\hat{h}_1)-\mu_2(x|\hat{h}_1)+\mu_2(x|\hat{h}_1)-\lambda(x)}{\lambda(x)}\right]^{\frac{1}{2}} =$$

$$\lambda(x)^{\frac{1}{2}}\left[1+\frac{\hat{\lambda}(x|\hat{h}_1)}{\lambda(x)}-1\right]^{\frac{1}{2}} = \lambda(x)^{\frac{1}{2}}\frac{\hat{\lambda}(x|\hat{h}_1)^{\frac{1}{2}}}{\lambda(x)^{\frac{1}{2}}} = \hat{\lambda}(x|\hat{h}_1)^{\frac{1}{2}}.$$

Lemma A.2 (Equation (27)).

In order to show that (27) is true we only need to show that

$$\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda(X_i)^{\frac{1}{2}}\delta(X_i)L\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}+\delta_3(x)$$
$$=\frac{1}{2}\varepsilon_1(x|\hat{h}_1,\hat{h}_2)+\frac{1}{2}\varepsilon_2(x|\hat{h}_1,\hat{h}_2)+\varepsilon_3(x|\hat{h}_1,\hat{h}_2).$$

Multiply and divide the numerator and the denominator of both ε_1 and ε_2 by $\lambda(X_i)$. Then,

$$\frac{\varepsilon_1(x|\hat{h}_1,\hat{h}_2)+\varepsilon_2(x|\hat{h}_1,\hat{h}_2)}{2}=\frac{1}{n\hat{h}_2}\sum_{i=1}^{n}\frac{\lambda^{\frac{1}{2}}(X_i)\delta_4(X_i)L\left\{\left(\frac{x-X_i}{\hat{h}_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}.$$

Since

$$\sup_{x\in[0,T]}\{\mathrm{D}(x|\hat{h}_1)^2+b(x|\hat{h}_1)^2\}\le\sup_{x\in[0,T]}\left\{\mathrm{D}(x|\hat{h}_1)+b(x|\hat{h}_1)\right\}^2,$$

taking

$$\varepsilon_4(x|\hat{h}_1,\hat{h}_2)=\frac{1}{n\hat{h}_2}\left\{\sup_{x\in[0,T]}\{\mathrm{D}(x|\hat{h}_1)^2+b(x|\hat{h}_1)^2\}\sum_{i=1}^{n}\mathrm{I}\left(|x-X_i|\le C_2\hat{h}_2\right)\right\}$$

(27) follows immediately.

Lemma A.3 (Equivalence of (28) with (29) and (30)).

Applying the definition of convergence in probability to (28) we see that we need to show that

$$\mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{\varepsilon_1(x|h_1,h_2)+\varepsilon_2(x|h_1,h_2)}{2}+\varepsilon_3(x|h_1,h_2)-T(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}\xrightarrow{p}0.$$

Now,

$$\mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{\varepsilon_1(x|h_1,h_2)+\varepsilon_2(x|h_1,h_2)}{2}+\varepsilon_3(x|h_1,h_2)-T(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}$$

$$\leq \mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_1(x|h_1,h_2)-T(x|h_1,h_2)\right|>\frac{\xi n^{-\frac{4}{9}}}{3}\right\}+$$

$$+\mathrm{P}\left\{\sup_{x\in[0,T]}\left|\frac{1}{2}\varepsilon_2(x|h_1,h_2)\right|>\frac{\xi n^{-\frac{4}{9}}}{3}\right\}+\mathrm{P}\left\{\sup_{x\in[0,T]}|\varepsilon_3(x|h_1,h_2)|>\frac{\xi n^{-\frac{4}{9}}}{3}\right\}$$

In order to show that every term in the above sum is $O\left(n^{-r}\right)$, for every n we consider a finite subset $\mathcal{X}$ of $(0,T)$. Then for all $x\in\mathcal{X}$ there exists $y\in[0,T]$ such that $|x-y|<n^{-s}$, for an arbitrary positive s. Since the functions ε_i, $i=1,2,4$ are Hölder continuous, as sums of Hölder continuous functions, equivalently we can show that for any $\xi,r>0$

$$\sup_{x\in[0,T]}\mathrm{P}\left\{\left|\frac{1}{2}\varepsilon_1(x|h_1,h_2)-T(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}=O\left(n^{-r}\right)$$

$$\sup_{x\in[0,T]}\mathrm{P}\left\{\left|\frac{1}{2}\varepsilon_i(x|h_1,h_2)\right|>\xi n^{-\frac{4}{9}}\right\}=O\left(n^{-r}\right),\quad i=2,4.$$

Lemma A.4 (Bound for $|\mathrm{D}(x|h_1)|$)**.**

Define the sequence of σ-fields $\mathcal{F}_n=\sigma(X_1,X_2,\dots,X_n)$. Set

$$\lambda_i(x)=\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}-\mathbb{E}\left\{\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}.$$

For every $i=1,2,\dots,n$

$$\mathbb{E}\lambda_i(x)=\mathbb{E}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}-\mathbb{E}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}=0$$

and

$$\mathbb{E}\left(\sum_{i=1}^{n}\lambda_i(x)|\mathcal{F}_{n-1}\right)=\mathbb{E}\left(\sum_{i=1}^{n-1}\lambda_i(x)|\mathcal{F}_{n-1}\right)+\mathbb{E}\left(\lambda_n(x)|\mathcal{F}_{n-1}\right)=\sum_{i=1}^{n-1}\lambda_i(x),$$

i.e. $\sum_{i=1}^{n}\lambda_i(x)$ is a martingale with respect to the sequence of σ-fields generated by $\{\mathcal{F}_n,n\geq 1\}$. Since $\mathrm{D}(x|h_1)-\bar{\mathrm{D}}(x|h_1)\to 0$ a.s. as $n\to+\infty$, where

$$\bar{\mathrm{D}}(x|h_1)=\bar{\lambda}(x|h_1)-\mathbb{E}\left\{\bar{\lambda}(x|h_1)\right\}$$

we replace D by $\bar{\mathrm{D}}$.

$$\bar{\mathrm{D}}(x|h_1)=\bar{\lambda}(x|h_1)-\mathbb{E}\bar{\lambda}(x|h_1)=$$

$$\frac{1}{nh_1}\sum_{i=1}^{n}\left\{\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}-\mathbb{E}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}=\frac{1}{nh_1}\sum_{i=1}^{n}\lambda_i(x).$$

Now,

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2 \le \mathbb{E}\sum_{i=1}^{n}\left\{\frac{1}{nh_1}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}^2$$

$$= \frac{1}{nh_1^2}\int K^2\left(\frac{x-y}{h_1}\right)\frac{\lambda(y)}{1-F(y)}\,dy.$$

Set $x-y=h_1z$. Then,

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2 \le \frac{1}{nh_1}\int K^2(z)\frac{\lambda(x-h_1z)}{1-F(x-h_1z)}\,dz.$$

Approximating $\lambda(x-h_1z)/(1-F(x-h_1z))$ with its Taylor expansion around x, and since by conditions A1 and A3 $R(K)$ is finite, there is a constant $C_1>0$ such that

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2 \le \frac{C_1}{nh_1}.$$

Also,

$$\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^l \le \mathbb{E}\left\{\frac{1}{nh_1}\frac{K\left(\frac{x-X_i}{h_1}\right)}{1-F(X_i)}\right\}^l = \frac{1}{n^{l-1}h_1^l}\int K^l\left(\frac{x-y}{h_1}\right)\frac{\lambda(y)}{(1-F(y))^l}\,dy.$$

Using the same change of variable and Taylor expansion as above, there is a constant $C_2>0$ such that

$$\mathbb{E}\sum_{i=1}^{n}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^l \le \frac{C_2}{(nh_1)^{l-1}}.$$

Then, a bound for $\mathbb{E}|\mathrm{D}(x|h_1)|^l$ can be found by (31) with

$$\Phi(x)=x^l,\ f^*=\sup_{i=1,\ldots,n}\sum_{i=1}^{n}\frac{\lambda_i(x)}{nh_1},\ d_i=\frac{\lambda_i(x)}{nh_1},\ \text{and } S(f)=\left(\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2\right)^{\frac{1}{2}}.$$

Applying (31) yields,

$$\mathbb{E}\left|\sup_{i=1,\ldots,n}\sum_{i=1}^{n}\frac{\lambda_i(x)}{nh_1}\right|^l \le C\left[\sum_{i=1}^{n}\mathbb{E}\left\{\frac{\lambda_i(x)}{nh_1}\right\}^2\right]^{\frac{l}{2}} + c\sum_{i=1}^{n}\mathbb{E}\left|\frac{\lambda_i(x)}{nh_1}\right|^l.$$

Since,

$$\mathbb{E}|\bar{\mathrm{D}}(x|h_1)|^l \le \mathbb{E}\left|\sup_{i=1,\ldots,n}\sum_{i=1}^{n}\frac{\lambda_i(x)}{nh_1}\right|^l$$

we finally get that

$$\mathbb{E}|\bar{\mathrm{D}}(x|h_1)|^l \le C_3\left\{\left(\frac{1}{nh_1}\right)^{\frac{l}{2}} + \frac{1}{(nh_1)^{l-1}}\right\}.$$

Lemma A.5 (bound for $\varepsilon_2(x|h_1,h_2)$)**.**

$$\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} = \int \lambda^{\frac{1}{2}}(x-h_2z)b(x-h_2z|h_1)L\left\{z\lambda^{\frac{1}{2}}(x-h_2z)\right\}\,dz. \tag{47}$$

Using (3), (4) and $\eta = h_2\lambda^{-\frac{1}{2}}(x)$ yields $u(h_2z) = u\left(h_2y\lambda^{-\frac{1}{2}}(x)\right) = u(\eta y)$. Then,

$$\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} =$$

$$\int \frac{\lambda^{\frac{1}{2}}(x)}{\lambda^{\frac{1}{2}}(x)}\lambda^{\frac{1}{2}}(x-h_2z)b(x-h_2z|h_1)L\left\{z\frac{\lambda^{\frac{1}{2}}(x)}{\lambda^{\frac{1}{2}}(x)}\lambda^{\frac{1}{2}}(x-h_2z)\right\}\,dz$$

$$= \int u(\eta y)b(\eta y|h_1)L\left\{yu(\eta y)\right\}\,dy. \tag{48}$$

Expanding b in Taylor series around 0 and working as in (8)–(11) gives

$$\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} =$$

$$\int \left\{L(y) + \eta\left(L(y)u'(0)y + yb'(0)L(y) + y^2u'(0)L'(y)\right)\right\}\,dy + o(\eta^2).$$

Note that

$$\int L(y) = \int K(y)\,dy + yK(y) - \int K(y)\,dy = 0 \tag{49}$$

$$\int yL(y) = \int yK(y)\,dy + \int y^2K'(y)\,dy = 0. \tag{50}$$

Rearranging, and since $b(x|h_1) \leq C_1h_1^2$ and $\lambda(x-h_2z)/\lambda(x) \leq C_2h_2^2$ we get finally that

$$\mathbb{E}\left\{\varepsilon_2(x|h_1,h_2)\right\} \leq Mh_1^2h_2^2$$

where M is a positive generic constant.

Lemma A.6 (Bound for $\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)|$)**.**

Set

$$r_i(x) = \frac{1}{nh_2}\frac{\lambda^{-\frac{1}{2}}(X_i)b(X_i|h_1)L\left\{\left(\frac{x-X_i}{h_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)}.$$

Then

$$\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| = \mathbb{E}\left|\sum_{i=1}^n r_i(x) - \mathbb{E}\sum_{i=1}^n r_i(x)\right| = \mathbb{E}\left|\sum_{i=1}^n\left\{r_i(x) - \mathbb{E}r_i(x)\right\}\right|.$$

Now, let

$$Z_i = r_i(x) - \mathbb{E}r_i(x),\quad i = 1,\dots,n.$$

For every $i = 1,2,\dots,n$

$$\mathbb{E}Z_i = \mathbb{E}r_i(x) - \mathbb{E}r_i(x) = 0.$$

Also,

$$\mathbb{E}\left(\sum_{i=1}^n Z_i|\mathcal{F}_{n-1}\right) = \mathbb{E}\left(\sum_{i=1}^{n-1} Z_i|\mathcal{F}_{n-1}\right) + \mathbb{E}\left(Z_n|\mathcal{F}_{n-1}\right) = \sum_{i=1}^{n-1} Z_i,$$

i.e. $\sum_{i=1}^{n} Z_i$ is a martingale with respect to the sequence of σ-fields generated by $\{\mathcal{F}_n, n \geq 1\}$. Now,

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}(Z_i)^2 &\leq \sum_{i=1}^{n} \mathbb{E}(r_i(x))^2 \\ &= \int \sum_{i=1}^{n} \frac{1}{n^2 h_2^2} \frac{\lambda(u) b^2(u|h_1)}{(1-F(u))^2} L^2 \left(\frac{x-u}{h_2} \lambda^{\frac{1}{2}}(u) \right) f(u)\, du \\ &= \int \frac{1}{n h_2^2} \frac{\lambda^2(u)}{1-F(u)} b^2(u|h_1) L^2 \left(\frac{x-u}{h_2} \lambda^{\frac{1}{2}}(u) \right) du.\end{aligned}$$

Since $|b(x|h_1)| \leq C_6 h_1^2$, we have that $b(x|h_1)^2 \leq C_6^2 h_1^4$. Also, from conditions A1, A3, $R(K + zK')$ is finite. Thus, using changes of variables similar to those used in the proof of theorem 2.1 for the variance,

$$\sum_{i=1}^{n} \mathbb{E}(Z_i)^2 \leq C_7 \frac{h_1^4}{n h_2}.$$

Also,

$$\sum_{i=1}^{n} \mathbb{E}\{Z_i\}^l \leq \sum_{i=1}^{n} \int \frac{1}{n^l h_2^l} \frac{\lambda^{\frac{l}{2}}(u) b^l(u|h_1)}{(1-F(u))^l} L^l \left(\frac{x-u}{h_2} \lambda^{\frac{1}{2}}(u) \right) f(u)\, du \leq \frac{C_7 h_1^{2l}}{(n h_2)^{l-1}}.$$

Using (31) with

$$\Phi(x) = x^l,\ f^* = \sup_{i=1,\ldots,n} \sum_{i=1}^{n} Z_i,\ d_i = Z_i,\ \text{and}\ S(f) = \left(\sum_{i=1}^{n} \mathbb{E}\left\{Z_i\right\}^2 \right)^{\frac{1}{2}}$$

gives

$$\mathbb{E} \left| \sup_{i=1,\ldots,n} \sum_{i=1}^{n} Z_i \right|^l \leq C \left[\sum_{i=1}^{n} \mathbb{E}\left\{Z_i\right\}^2 \right]^{\frac{l}{2}}$$

and therefore

$$\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| \leq C_8 \left\{ \frac{h_1^{2l}}{(n h_2)^{\frac{l}{2}}} + \frac{h_1^{2l}}{(n h_2)^{l-1}} \right\} = C_8 \frac{h_1^{2l}(n h_2)^{l-1} + h_1^{2l}(n h_2)^{\frac{l}{2}}}{(n h_2)^{\frac{3l}{2}-1}}.$$

Recall that $h_2 \geq \eta n^{-1/9}$.

$$\frac{1}{h_2} \leq \frac{1}{\eta n^{-\frac{l}{9}}} \Rightarrow \frac{1}{n h_2} \leq \frac{1}{\eta n^{\frac{8}{9}}} \Rightarrow \left(\frac{1}{n h_2} \right)^{\frac{3l}{2}-1} \leq \frac{1}{\eta^{\frac{3l}{2}-1} n^{\frac{4l}{3}-\frac{8}{9}}}$$

Also

$$h_1^{2l} \leq \rho n^{2al - \frac{2l}{5}},\ \ (n h_2)^{l-1} \leq n^{-\frac{l}{9}}$$

and therefore

$$h_1^{2l}(n h_2)^{l-1} \leq C n^{-\frac{l}{9}} n^{2al - \frac{2l}{5}} = C n^{2al - \frac{19l}{45}}$$

Similarly

$$h_1^{2l}(n h_2)^{\frac{l}{2}} \leq n^{-\frac{l}{9}} \left(n \rho n^{a - \frac{2}{5}} \right)^{\frac{l}{2}}.$$

Hence,

$$\mathbb{E}|\varepsilon_2(x) - \mathbb{E}\varepsilon_2(x)| \le C\frac{n^{2al-\frac{19l}{45}} + n^{al+\frac{3l}{10}-\frac{l}{9}}}{\eta^{\frac{3l}{2}-1}n^{\frac{4l}{3}}}.$$

Lemma A.7 (Equation (35))**.**

From the definition of the function $M(x,y)$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} M(X_{(i)},X_{(j)}) = \sum\sum_{i\neq j} M(X_i,X_j) + \sum_{i=1}^{n} M(X_i,X_i).$$

Thus,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} M(X_{(i)},X_{(j)}) =$$

$$\sum\sum_{i\neq j}\{m(X_i,X_j) - m_1(X_i)\} + \sum_{i=1}^{n}\{m(X_i,X_i) - m_1(X_i)\} =$$

$$\sum\sum_{i\neq j} m(X_i,X_j) + \sum_{i=1}^{n} m(X_i,X_i) - \left((n-1)\sum_{i=1}^{n} m_1(X_i) + \sum_{i=1}^{n} m_1(X_i)\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} m(X_i,X_j) - n\sum_{i=1}^{n} m_1(X_i).$$

Therefore,

$$\frac{1}{n^2h_1h_2}\left\{\sum\sum_{i\neq j} M(X_i,X_j) + \sum_{i=1}^{n} M(X_i,X_i)\right\} + 2T(x|h_1,h_2) =$$

$$\frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} m(X_i,X_j) - \frac{1}{nh_1h_2}\sum_{i=1}^{n} m_1(X_i) + \frac{1}{nh_1h_2}\sum_{i=1}^{n} m_1(X_i)$$

$$= \frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} m(X_i,X_j).$$

Now,

$$\frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n} m(X_i,X_j) =$$

$$\frac{1}{n^2h_1h_2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\lambda(X_j)^{\frac{1}{2}}h_1\left\{\frac{K_{h_1}(X_j-X_i)}{1-F(X_j)} - \mu(X_j|h_1)\right\}\mathrm{L}\left\{\frac{x-X_j}{h_2}\lambda(X_j)^{\frac{1}{2}}\right\}}{1-F(X_j)} =$$

$$\frac{1}{nh_2}\sum_{i=1}^{n}\frac{\lambda^{-\frac{1}{2}}(X_i)\mathrm{D}(X_i|h_1)L\left\{\left(\frac{x-X_i}{h_2}\right)\lambda(X_i)^{\frac{1}{2}}\right\}}{1-F(X_i)} = \varepsilon_1(x|h_1,h_2).$$

References

[1] Abramson, I. (1982). Arbitrariness of the pilot estimator in adaptive kernel methods, *Journal of Multivariate Analysis*, **12**, 562–567.

[2] Bagkavos, D. (2003). *Bias reduction in nonparametric hazard rate estimation. PhD thesis*, The University of Birmingham, UK.

[3] Burkholder, D.L. (1973). Distribution function inequalities for martingales, *The Annals of Probability*, **1**, 19–42.

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

[5] Hall, P. (1990). On the bias of variable bandwidth curve estimators, *Biometrika*, **77**, 529–535.

[6] Hall, P. and Heyde, C. (1980). *Martingale limit theory and its applications*, Academic Press, New York.

[7] Hall, P. and Marron, J.S. (1988). Variable window width kernel estimates of probability densities, *Probability Theory and Related Fields*, **80**, 37–49.

[8] Hall, P., Hu, T.C. and Marron, J. S.(1995). Improved variable window width kernel estimates of probability densities, *Annals of Statistics*, **23**, 1–10.

[9] Hajek, J. (1968). Asymptotic normality of simple linear statistics under alternatives, *Annals of Mathematical Statistics*, **39**, 325–346.

[10] Lo, Y. Mack, S. and Wang, J. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meyer estimator., *Probability Theory and Related Fields*, **80**, 461–473.

[11] Patil, P.N. (1993). Bandwidth choice for non parametric hazard rate estimation, *Journal of Statistical Planning and Inference*, **35**, 15–30.

[12] Rice, J. and Rosenblatt, M. (1976). Estimation of the log-survivor function and hazard function, *Sankhya, ser. a*, **38**, 60–78.

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

[14] Sarda, P. and Vieu, P. (1991). Smoothing Parameter Selection In Hazard Rate Estimation, *Statistics and Probability Letters*, **11**, 429–434.

[15] Silverman, B. (1986). *Density estimation for statistics and data analysis*, Chapman & Hall, London.

[16] Stone, C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates, *The Annals of Statistics*, **12**, 1285–1297.

[17] Tanner, M. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data, *The Annals of Statistics*, **11**, 994–998.

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

[19] Watson, G. and Leadbetter, M. (1964). Hazard analysis I, *Biometrika*, **51**, 175–184.

[20] Stone, C.J. (1984) An Asymptotically Optimal Window Selection Rule For Kernel Density Estimates, *Annals of Statistics***12, (4)**, 1285–1297

School of Mathematics and Statitstics, The University of Birmingham, Edgbaston, Birmingham, B152TT, UK

E-mail address: `dimitrios.bagkavos@accenture.com`

E-mail address: `dimitrios.bagkavos@gmail.com`