Adaptive Kernel Estimation of The Hazard Rate Function

Raid Salha
Department of Mathematics,
Islamic University of Gaza, Palestine,
e-mail: rbsalha@mail.iugaza.edu

Abstract

In this paper, we generalized the constant bandwidth kernel estimator of the hazard rate function from Watson and Leadbetter (1964), which depends on a single bandwidth to the adaptive kernel estimator, which depends on different bandwidths. We derive the asymptotic normality of the adaptive kernel estimator. Also we illustrate the use of the adaptive kernel hazard rate estimator in both simulation and real life data and compared it to the constant bandwidth kernel estimator. In our study, we show that the adaptive estimator has no boundary effects as the constant bandwidth kernel estimator, and has smaller bias.

Keywords: *Hazard rate function, adaptive kernel estimation, constant bandwidth kernel estimation, density estimation.*

1 Introduction

Hazard rate function is important, since it provides information in reliability theory and survival analysis as well as in many fields as engineering, medical statistics and geophysics.

The hazard rate function $r(\cdot)$ is defined as

$$r(x) = \lim_{\triangle x \longrightarrow 0} \frac{P(X \leq x + \triangle x | X > x)}{\triangle x}, \quad x > 0,$$

and it can be written as

$$r(x) = \frac{f(x)}{1 - F(x)},$$

where $f(\cdot)$ and $F(\cdot)$ are the density and distribution function of a continuous random variable X respectively.

If $X_1, X_2, \ldots, X_n$ is a random sample distributed as X, then Watson and leadbetters (1964) proposed the following estimator for $r(\cdot)$

$$\hat{r}(x) = \frac{\hat{f}(x)}{1 - \hat{F}(x)}, \tag{1}$$

where $\hat{f}(x) = \dfrac{1}{nh} \displaystyle\sum_{i=1}^{n} K(\frac{x - X_i}{h})$, and $\hat{F}(x) = \dfrac{1}{nh} \displaystyle\sum_{i=1}^{n} \int_{-\infty}^{x} K(\frac{u - X_i}{h}) du$, where K is a bounded and symmetric kernel, integrating to one.

The estimator in (1) depends on a single bandwidth h, and we will call it the constant bandwidth kernel estimator in this paper. An extension of (1) to an adaptive kernel estimation that having different n bandwidth depend on X_i, $i = 1, 2, \ldots, n$ is important since the bandwidth plays an important role as a smoothing parameter. The basic idea of the adaptive kernel estimation is to construct a kernel estimator consisting of kernels placed at the observed data points, but allows the bandwidth of kernels to vary from point to another.

The adaptive kernel estimator $r_n(x)$ of $r(x)$ is defined as

$$r_n(x) = \frac{f_n(x)}{1 - F_n(x)}, \tag{2}$$

where

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i h}K(\frac{x - X_i}{\lambda_i h}),$$

and

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i h}\int_{-\infty}^{x} K(\frac{u - X_i}{\lambda_i h})\, du,$$

are adaptive kernel estimators of $f(x)$ and $F(x)$ respectively.
$\lambda_i = \{\frac{\widetilde{f}(X_i)}{g}\}^{-\alpha}$, where $\widetilde{f}(x)$ is a pilot estimate that satisfies $\widetilde{f}(X_i) > 0$ for all i and $g = \{\prod_{i=1}^{n}\widetilde{f}(X_i)\}^{\frac{1}{n}}$, $0 \leq \alpha \leq 1$.
Abramson (1982) shows that taking $\alpha = 0.5$ is a good choice since one can achieve a bias of order h^4 rather than h^2. For more details, see Silverman (1986), Wand and Jones (1995), and Fan and Gijbels (1992).

2 Conditions

The following conditions will be used in the sequel:

C1 Suppose that the kernel function K satisfies the following:

- (i) K is asymmetric density function.
- (ii) $\lim\limits_{y\longrightarrow\infty} |y|K(y) = 0$.
- (iii) $\int_{-\infty}^{\infty} K^2(u)du < \infty$.
- (iv) $\int_{-\infty}^{\infty} uK(u)du = 0$.
- (v) $\int_{-\infty}^{\infty} u^2K(u)du < \infty$.

C2 Suppose that the bandwidth h satisfies the following:

- (i) $h \longrightarrow 0$.
- (ii) $nh \longrightarrow \infty$.
- (iii) $nh^5 \longrightarrow 0$.

C3 f'' exists and integrable.

3 Preliminary Lemmas

In this section, we state and prove some basic lemmas that we need to prove our main results.

Lemma 1 (Bochner Lemma). Suppose that K satisfies the conditions **C1**(i), (ii) . Let $g(x)$ satisfy $\int_{-\infty}^{\infty} g(x)dx < \infty$. Let h be a sequence of positive constants satisfying **C2**(i). Then at every point x of continuity of $g(\cdot)$,

$$\lim_{n\longrightarrow\infty} \frac{1}{h}\int_{-\infty}^{\infty} K(\frac{y-x}{h})g(y)\,dy = g(x)\int_{-\infty}^{\infty} K(y)\,dy.$$

Proof. See Parzen (1962).

Lemma 2. Under the conditions **C1**(i), (ii), (iv), (v), **C2**(iii) and **C3**, then $(nh)^{\frac{1}{2}}|F_n(x) - F(x)|$ converges in probability to zero.
Proof.

$$\begin{aligned}(nh)^{\frac{1}{2}}|F_n(x) - F(x)| &= (nh)^{\frac{1}{2}}|F_n(x) - EF_n(x) + EF_n(x) - F(x)| \\ &\leq (nh)^{\frac{1}{2}}|F_n(x) - EF_n(x)| \\ &+ (nh)^{\frac{1}{2}}|EF_n(x) - F(x)|. \end{aligned} \tag{3}$$

First we will show that

$$\lim_{n\longrightarrow\infty}(nh)^{\frac{1}{2}}|EF_n(x) - F(x)| = 0. \tag{4}$$

$$\begin{aligned} EF_n(x) &= E\int_{-\infty}^{x} \frac{1}{\lambda h}K(\frac{u-X}{\lambda h})\,du \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} \frac{1}{\lambda h}K(\frac{u-s}{\lambda h})f(s)\,du\,ds \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} K(v)f(\lambda hv + u)\,dv\,du \\ &= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u)K(v)\,du\,dv \\ &- \int_{-\infty}^{x}\int_{-\infty}^{\infty} f'(u)\lambda hvK(v)\,dv\,du \\ &+ \int_{-\infty}^{x}\int_{-\infty}^{\infty} \frac{1}{2}f''(u)(\lambda h)^2v^2K(v)\,dv\,du + o(h^2) \\ &= F(x) + o(h^2). \end{aligned}$$

$|EF_n(x) - F(x)| = o(h^2)$. $(nh)^{\frac{1}{2}}|EF_n(x) - F(x)| = o((nh^5)^{\frac{1}{2}})$.
Using Condition **C2**(iii) completes the proof of (4).

Secondly we will prove that

$$\lim_{n \longrightarrow \infty} (nh)^{\frac{1}{2}} |F_n(x) - EF_n(x)| = 0. \tag{5}$$

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i h} \int_{-\infty}^{x} K(\frac{u - X_i}{\lambda_i h})\, du = \frac{1}{n} \sum_{i=1}^{n} W_i(x),$$

where

$$W_i(x) = \frac{1}{\lambda_i h} \int_{-\infty}^{x} K(\frac{u - X_i}{\lambda_i h})\, du.$$

Let $\varepsilon > 0,\ \delta > 0$ be given.

$$\begin{aligned}
P((nh)^{\frac{1}{2}}|F_n(x) - EF_n(x)| > \varepsilon) &\leq \varepsilon^{-2-2\delta}(nh)^{1+\delta} E|\frac{1}{n}\sum_{i=1}^{n}[W_i(x) - EW_i(x)]|^{2+2\delta} \\
&= \varepsilon^{-2-2\delta} h^{1+\delta} n^{-1-\delta} E|\sum_{i=1}^{n}[W_i(x) - EW_i(x)]|^{2+2\delta} \\
&\leq 2^{1+\delta}\varepsilon^{-2-2\delta}(n^{-1}h)^{1+\delta}\sum_{i=1}^{n} E|W_i(x)|^{2+2\delta} \\
&+ 2^{1+\delta}\varepsilon^{-2-2\delta}(n^{-1}h)^{1+\delta}\sum_{i=1}^{n} |EW_i(x)|^{2+2\delta}
\end{aligned}$$

$$\begin{aligned}
(n^{-1}h)^{1+\delta}\sum_{i=1}^{n} E|W_i(x)|^{2+2\delta} &= (n^{-1}h)^{1+\delta} n \int_{-\infty}^{\infty}\int_{-\infty}^{x} |\frac{1}{\lambda h} K(\frac{u - v}{\lambda h})|^{2+2\delta} f(v)\, du\, dv \\
&= n^{-\delta} h^{1+\delta} (\lambda h)^{-(1+2\delta)} \int_{-\infty}^{\infty}\int_{-\infty}^{x} |K(s)|^{2+2\delta} f(u - \lambda h s)\, du\, ds \\
&\leq C(nh)^{-\delta} \longrightarrow 0.
\end{aligned}$$

Similarly, $(n^{-1}h)^{1+\delta}\sum_{i=1}^{n} |EW_i(x)|^{2+2\delta} \longrightarrow 0$. This completes the proof of Equation (5).
Now substitution of the Equations (4) and (5) in Equation (3) completes the proof of the lemma. $\square$

Lemma 3. Under the conditions **C1**(i), (ii), (iii) and **C2**(i), we have

$$Var(f_n(x)) \simeq \frac{1}{nhg^{\alpha}} f(x)^{1+\alpha} \int_{-\infty}^{\infty} K^2(u)du.$$

Proof.

$$Var(f_n(x)) = \frac{1}{n} Var\left(\frac{1}{\lambda h} K(\frac{x - X_i}{\lambda h})\right).$$

let $\widetilde{h} = \lambda h$. Since $h \longrightarrow 0$, $\widetilde{h} \longrightarrow 0$.

$$\begin{aligned} nhVar(f_n(x)) &= hVar\left(\frac{1}{\widetilde{h}} K(\frac{x - X_i}{\widetilde{h}})\right) = hE\left(\frac{1}{\widetilde{h}^2} K^2(\frac{x - X_i}{\widetilde{h}})\right) \\ &= h\int_{-\infty}^{\infty} \frac{1}{\widetilde{h}} K^2(\frac{x-u}{\widetilde{h}}) f(u) \frac{1}{hg^{\alpha} \widetilde{f}(u)^{-\alpha}} du. \end{aligned}$$

Since $\widetilde{f}(u)$ is an estimate of $f(u)$, then $\widetilde{f}(u) \longrightarrow f(u)$ as $n \longrightarrow \infty$ and by an application of Bochner lemma the proof of the lemma is completed. □

4 Main Results

In this section, we state and prove our main results.
Theorem 1. Under the conditions **C1**(i), (ii), (iii) and **C2**(i), (ii), we have

$$(nh)^{\frac{1}{2}}(f_n(x) - f(x)) \xrightarrow{\mathrm{D}} N(0, \frac{1}{hg^{\alpha}} f(x)^{1+\alpha} \int_{-\infty}^{\infty} K^2(u)du).$$

Proof.
Let

$$V_{ni} = \frac{1}{\lambda_i h} K(\frac{x - X_i}{\lambda_i h}).$$

Then $f_n(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} V_{ni}$, where $V_{ni}, i = 1, 2, \ldots, n$ are iid random variables as $V_n = \dfrac{1}{\lambda h} K(\dfrac{x - X}{\lambda h})$.
Now we want to show that Liapounov condition is satisfied, that is for some $\delta > 0$

$$\lim_{n \longrightarrow \infty} \frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\delta/2} \sigma^{2+\delta}(V_n)} = 0,$$

see Pranab and Julio (1993).

$$E|V_n|^{2+\delta} = \int_{-\infty}^{\infty} |\frac{1}{\lambda h} K(\frac{x-y}{\lambda h})|^{2+\delta} f(y)dy \longrightarrow \frac{1}{(\lambda h)^{1+\delta}} f(x) \int_{-\infty}^{\infty} |K(y)|^{2+\delta} dy.$$

$$Var(V_n(x)) \simeq \frac{1}{hg^{\alpha}} f(x)^{1+\alpha} \int_{-\infty}^{\infty} K^2(u)du.$$

$$\frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\delta/2}\sigma^{2+\delta}(V_n)} = \frac{(\lambda h)^{1+\delta} E|V_n - E[V_n]|^{2+\delta}}{\lambda^{1+\delta}(nh)^{\delta/2} h^{1+\delta/2}\sigma^{2+\delta}(V_n)} \longrightarrow 0,$$

since

$$(\lambda h)^{1+\delta} E|V_n|^{2+\delta} \longrightarrow f(x) \int_{-\infty}^{\infty} |K(u)|^{2+\delta} du < \infty,$$

$$h^{1+\delta/2}\sigma^{2+\delta}(V_n) = (h\sigma^2(V_n))^{\frac{2+\delta}{2}} \longrightarrow (g^{-\alpha} f(x)^{1+\alpha} \int_{-\infty}^{\infty} K^2(u)du)^{\frac{2+\delta}{2}} < \infty,$$

and $(nh)^{\delta/2} \longrightarrow \infty$, by condition **C2**(ii).
This implies that $\{f_n(x)\}$ is asymptotically normally distributed with mean $f(x)$ and variance $\frac{1}{hg^{\alpha}} f(x)^{1+\alpha} \int_{-\infty}^{\infty} K^2(u)du$. This completes the proof of the Theorem. □

Theorem 2. Under the conditions **C1**, **C2** and **C3** the following is true

$$(nh)^{\frac{1}{2}}(r_n(x) - r(x)) \xrightarrow{\mathrm{D}} N(0, \frac{1}{g^{\alpha}} \frac{z^2(x)}{f(x)^{1-\alpha}} \int_{-\infty}^{\infty} K^2(u)du).$$

Proof.

$$\begin{aligned}
(nh)^{\frac{1}{2}}(r_n(x) - r(x)) &= (nh)^{\frac{1}{2}} \left[\frac{f_n(x)}{1-F_n(x)} - \frac{f(x)}{1-F(x)} \right] \\
&= (nh)^{\frac{1}{2}} \left[\frac{f_n(x)}{1-F_n(x)} - \frac{f(x)}{1-F_n(x)} - \frac{f(x)}{1-F(x)} + \frac{f(x)}{1-F_n(x)} \right] \\
&= \frac{(nh)^{\frac{1}{2}}}{1-F_n(x)} \left[f_n(x) - f(x) - \frac{f(x)(1-F_n(x))}{1-F(x)} + f(x) \right] \\
&= \frac{(nh)^{\frac{1}{2}}}{1-F_n(x)} \left[f_n(x) - f(x) \right] \\
&+ \frac{(nh)^{\frac{1}{2}} f(x)}{(1-F_n(x))(1-F(x))} \left[F_n(x) - F(x) \right].
\end{aligned}$$

By Lemma 2 and Theorem 1, the proof is completed. □

5 Applications

We now illustrate the method of adaptive kernel estimator, r_n, via simulation study and a real data set. The purpose of the applications is to demonstrate that the adaptive method works reasonably well especially in the tails and it is better than the constant bandwidth kernel estimator, $\hat{r}$. Throughout this section, the Epanechnikov kernel $K(x) = 0.75(1 - x^2)I_{|x|<1}$, where I denotes the indicator function is used.

5.1 A simulation Application

We simulated a data of size 200 from a standard normal $N(0, 1)$. After that we evaluated the hazard rate function of the standard normal at 61 points in the interval [0,3]. We estimated the standard normal hazard rate function at the same points by using the estimators r_n and $\hat{r}$. Figure 1 shows that the adaptive is reasonable good and performs better than the constant bandwidth especially at the tails of the hazard rate function. Figure 2 gives estimated standard deviations of the two estimators appears in , from which we see that the two standard deviations are very close in the interior but in the tail the standard deviation of $\hat{r}$ is very large compared to the standard deviation of r_n. Finally, we calculated a 95% confidence intervals using Theorem 2 for the first and last 5 observations. We found that the confidence intervals contain all the corresponding true values and the average lengths of the first 5 confidence intervals, the last 5 confidence intervals, and the average length of the 10 confidence intervals were 28.8%, 85.1%, and 57.0% of the range of the data, respectively.

i	True value	95% conf. int.	i	True value	95% conf. int.
1	0.799	(0.357, 1.131)	57	3.099	(1.559, 3.788)
2	0.830	(0.361, 1.141)	58	3.1440	(1.590, 3.875)
3	0.863	(0.376, 1.168)	59	3.190	(1.566, 3.889)
4	0.896	(0.410, 1.222)	60	3.237	(1.699, 4.121]
5	0.929	(0.461, 1.302)	61	3.283	(1.888, 4.435)

Table 1: 95% confidence intervals for the standard normal hazard rate function

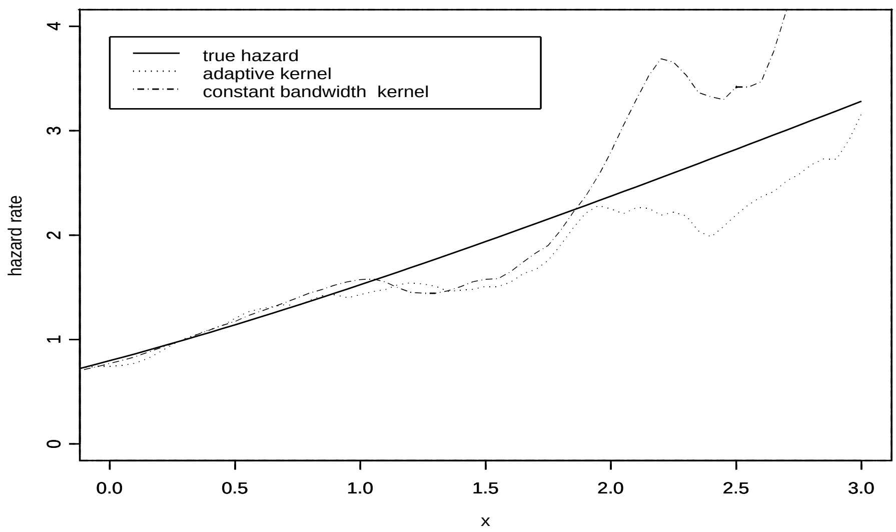

Figure 1: True hazard rate function, adaptive and constant kernel hazard rate estimates

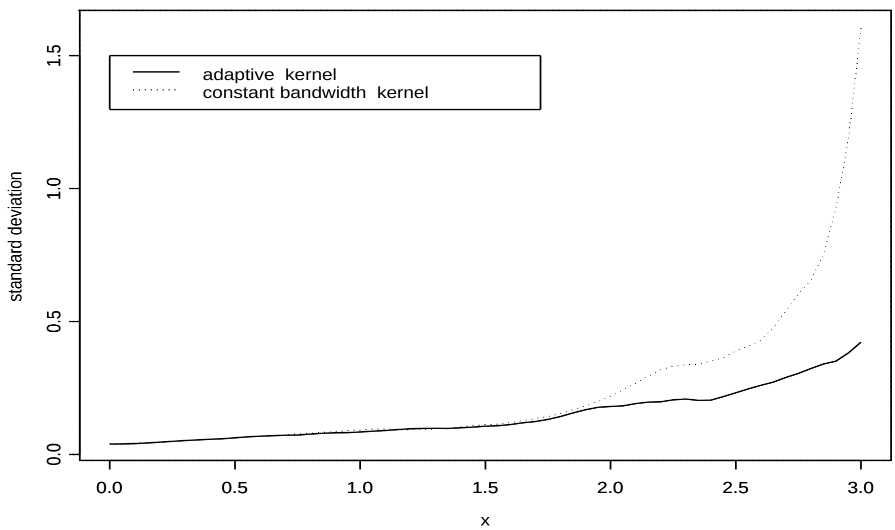

Figure 2: The estimated standard deviations of the adaptive and the constant bandwidth kernel hazard rate estimates

5.2 Application for a real life data

In this subsection, we use the suicide data given in Silverman (1986: pp.8), to exhibit the practical performance of the adaptive kernel estimator r_n. The data gives the lengths of the treatment spells in days of control patients in a suicide study. We estimate the hazard rate which represents the instant potential per unit of time that an individual commits suicide within the time interval $(t, t + \Delta t)$ given that it was known to alive up to time t.

For comparison purposes we also estimate the hazard rate function using the constant bandwidth kernel estimator $\hat{r}$. For both estimators we used the Epanechnikov kernel. The result is given in Figure 3. Although the suggested value of the hazard rate function from the two estimators is different, they both suggest a similar structure for the hazard rate function. As we see the behavior of the two estimators is very similar in the interior especially from approximately $x = 300$ to $x = 450$, and the divergence of the two estimators gets large at the boundary.

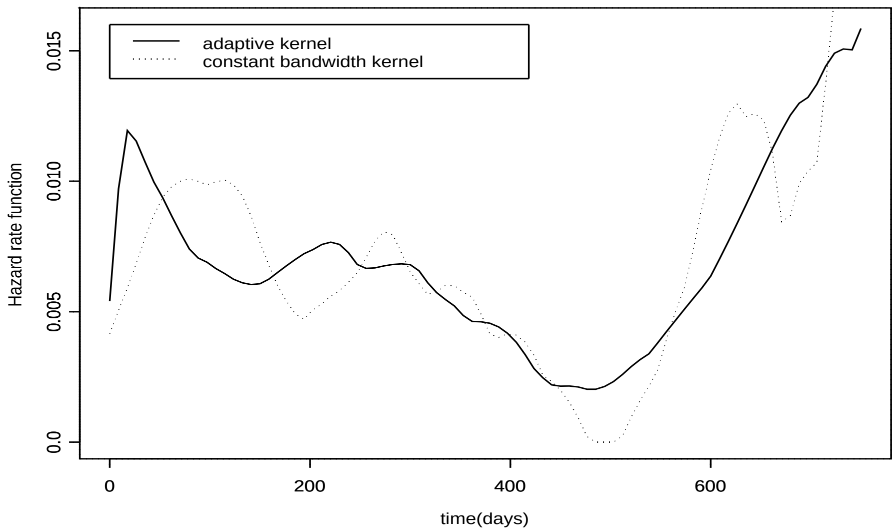

Figure 3: The adaptive and the constant bandwidth kernel hazard rate estimates for the suicide data

References

[1] Abramson, I. (1982). On the bandwidth variation in kernel estimation-a square root law. The Annals of Statistics, 10, 1217-1223.

[2] Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. The Annals of Statistics, 20, 2008-2036.

[3] Parzen, E., (1962). On estimation of a probability density function and mode. Annals of Mathematical Statistics, Vol. 33,1065-1076.

[4] Pranab K. and Julio M. (1993). Large sample methods in statistics. Chapman & Hall. Inc.

[5] Silverman, B. W. (1986). Density estimation for statistics and data analysis.Chapman & Hall. Inc.

[6] Wand, M. P. and Jones, M. C. (1995), Kernel smoothing. Chapman & Hall. Inc.

[7] Watson, G. and Leadbetter, M. (1964). Hazard Analysis I. Biometrika, 51, 175-184.