# On variable bandwidth kernel density estimation

Janet Nakarmi Hailin Sang*

**Abstract**
In this paper we study the ideal variable bandwidth kernel estimator introduced by McKay [7, 8] and the plug-in practical version of variable bandwidth kernel estimator with two sequences of bandwidths as in Giné and Sang [4]. The dominating terms of the variance of the true estimator in the variance decomposition are separated from the other terms. Based on the exact formula of bias and these dominating terms, we develop the optimal bandwidth selection of this variable kernel estimator.

**Key Words:** bandwidth selection, variable kernel density estimation.

## 1. Introduction

Suppose that $X_i, i \in \mathbb{N}$, are independent identically distributed (i.i.d.) observations with density function $f(t)$, $t \in \mathbb{R}$. Let $K$ to be a symmetric probability kernel satisfying some differentiability properties. The classical kernel density estimator

$$\hat{f}(t; h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n}\right), \tag{1}$$

where $h_n$ is the bandwidth sequence with $h_n \to 0, nh_n \to \infty$, and its properties have been well studied in the literature. The variance of (1) has order $O(nh_n)^{-1}$ and the bias has order $h_n^2$ if $f(t)$ has bounded second order derivative. In this paper, we study the following variable bandwidth kernel density estimator proposed by McKay [7, 8]:

$$\bar{f}(t; h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} \alpha(f(X_i)) K(h_n^{-1} \alpha(f(X_i))(t - X_i)), \tag{2}$$

where $\alpha(s)$ is a smooth function of the form

$$\alpha(s) := cp^{1/2}(s/c^2). \tag{3}$$

The function $p$ is at least four times differentiable and satisfies $p(x) \geq 1$ for all $x$ and $p(x) = x$ for all $x \geq t_0$ for some $1 \leq t_0 < \infty$, and $0 < c < \infty$ is a fixed number. (2) is a variable bandwidth kernel density estimator since the bandwidth has form $h_n/\alpha(f(X_i))$ if we rewrite (2) in the form of the classical one, (1). The study of variable bandwidth kernel density estimation goes back to Abramson [1]. He proposed the following estimator

$$f_A(t; h_n) = n^{-1} \sum_{i=1}^{n} h_n^{-1} \gamma(t, X_i) K(h_n^{-1} \gamma(t, X_i)(t - X_i)), \tag{4}$$

*Department of Mathematics, University of Mississippi, University, MS 38677, USA

where, $\gamma(t,s) = (f(s) \vee f(t)/10)^{1/2}$. The bandwidth $h_n/\gamma(t, X_i)$ at each observation $X_i$ is inversely proportional to $f^{1/2}(X_i)$ if $f(X_i) \geq f(t)/10$. Notice that (2) also has the square root law since $\alpha(f(X_i)) = f^{1/2}(X_i)$ if $f(X_i) \geq t_0 c^2$ by the definition of the function $p(x)$. The estimator (2) or (4) has clipping procedure in $\alpha(s)$ or $\gamma(t,s)$ since they make the true bandwidth $h_n/\alpha(f(X_i)) \geq h_n/c$ or $h_n/\gamma(t, X_i) \geq 10^{1/2} h_n/f(t)^{1/2}$. The clipping procedure prevent too much contribution to the density estimation at $t$ if the observation $X_i$ is too far away from $t$. Abramson showed that, this square root law improves the bias from the order of $h_n^2$ to the order of $h_n^4$ for the estimator (4) while at the same time keeps the variance at the order of $(nh_n)^{-1}$ if $f(t) \neq 0$ and $f(x)$ has fourth order continuous derivative at $t$. However, this variable bandwidth estimator (4) is not a density function of a true probability measure since the integral of $f_A(t; h_n)$ over $t$ is not 1 -it would if $\gamma$ depended only on $s$-.

Terrell and Scott [11] and McKay [8] showed that the following modification of Abramson estimator without the 'clipping filter' $(f(t)/10)^{1/2}$ on $f^{1/2}(X_i)$ studied in Hall and Marron [5],

$$f_{HM}(t; h_n) = n^{-1} \sum_{i=1}^{n} h_n^{-1} f^{1/2}(X_i) K(h_n^{-1} f^{1/2}(X_i)(t - X_i)), \qquad (5)$$

which has integral 1 and hence is a true probability density, may have bias of order much larger than $h_n^4$. Therefore the clipping is necessary for such bias reduction.

Hall, Hu and Marron (1995) then proposed the estimator

$$f_{HHM}(t; h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n} f^{1/2}(X_i)\right) f^{1/2}(X_i) I(|t - X_i| < h_n B) \qquad (6)$$

where $B$ is a fixed constant; see also Novak [10] for a similar estimator. This estimator is non-negative and achieves the desired bias reduction but, like Abramson's, it does not integrate to 1.

In conclusion, it seems that the estimator (2) has all the advantages: it is a true density function with square root law and smooth clipping procedure. But notice that this estimator and all the other variable bandwidth kernel density estimators can not be applied in practice since they all include the studied density function $f$. Therefore, we call them ideal estimators in the literature. Hall and Marron [5] studied a true density estimator

$$\hat{f}_{HM}(t; h_{1,n}, h_{2,n}) = \frac{1}{nh_{2,n}} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_{2,n}} \hat{f}^{1/2}(X_i; h_{1,n})\right) \hat{f}^{1/2}(X_i; h_{1,n}),$$

by plugging in the classical estimator (1) as the pilot estimator. Here, the bandwidth sequence $h_{2,n}$ is the $h_n$ as in (5) and the bandwidth sequence $h_{1,n}$ is applied in the classical kernel density estimator (1). They took Taylor expansion of $K\left(\frac{t-X_i}{h_{2,n}} \hat{f}^{1/2}(X_i; h_{1,n})\right)$ at $K\left(\frac{t-X_i}{h_{2,n}} f^{1/2}(X_i)\right)$ and studied the asymptotics of the true estimator. By applying this Taylor decomposition, McKay [8] studied convergence of plug-in true estimator of (2) in probability and pointwise. Giné and Sang [3, 4] studied plug-in true estimators of (6) and (2) for one and multiple d-dimensional observations. They proved that the discrepancy between the true estimator and the true value converges uniformly over a data adaptive region at a rate $O((\log n/n)^{4/(8+d)})$ by applying empirical process techniques. The true estimator in Giné and Sang [4] has the following form

$$\hat{f}(t; h_{1,n}, h_{2,n}) = \frac{1}{nh_{2,n}} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_{2,n}} \alpha(\hat{f}(X_i; h_{1,n}))\right) \alpha(\hat{f}(X_i; h_{1,n})). \tag{7}$$

And, they also studied the uniform convergence in almost sure sense of true estimators with bias order $h_n^6$.

In section 2, we provide decomposition of some important terms which we use in later sections. In section 3 of this paper, we study the decomposition of the variance of the true estimator (7) and find the exact formula of the dominating terms in the variance decomposition. Moreover, we provide a theoretical formula for the optimal bandwidth in section 4.

## 2. Preliminary decomposition

For convenience, we adopt the notations as in Giné and Sang [4] for the Taylor series expansion of $K\left(\frac{t-X_i}{h_{2,n}}\alpha(\hat{f}(X_i; h_{1,n}))\right)$ at $K\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)$. We also give the statements without detailed explanation. For details, readers are referred to Giné and Sang [4] . $\mathcal{P}_C$ will denote the set of all probability densities on $\mathbb{R}$ that are uniformly continuous and are bounded by $C < \infty$, and $\mathcal{P}_{C,k}$ will denote the set of densities on $\mathbb{R}$ for which themselves and their partial derivatives of order $k$ or lower are bounded by $C < \infty$ and are uniformly continuous. Define $\delta(t) = \delta(t, n)$ by the equation

$$\delta(t) = \frac{\alpha(\hat{f}(t; h_{1,n})) - \alpha(f(t))}{\alpha(f(t))}.$$

Then

$$\alpha(\hat{f}(t; h_{1,n})) = \alpha(f(t))(1 + \delta(t)) \tag{8}$$

and

$$|\delta(t)| \leq Bc^{-2}|\hat{f}(t; h_{1,n}) - f(t)|$$

for a constant $B$ that depends only on $p$. Although we study the asymptotics of the true estimator pointwise, the uniform asymptotic behavior of the quantity $\delta(\cdot)$ is needed in the latter analysis. Define

$$D(t; h_{1,n}) = \hat{f}(t; h_{1,n}) - \mathbb{E}\hat{f}(t; h_{1,n}) \text{ and } b(t; h_{1,n}) = \mathbb{E}\hat{f}(t; h_{1,n}) - f(t).$$

Note that for $f \in \mathcal{P}_{C,2}$,

$$\sup_{t\in\mathbb{R}} |b(t; h_{1,n})| = O(h_{1,n}^2), \tag{9}$$

and by Giné and Guillou [2],

$$\sup_{t\in\mathbb{R}} |D(t; h_{1,n})| = O\left(\sqrt{\frac{\log h_{1,n}^{-1}}{nh_{1,n}}}\right)$$

for $f \in \mathcal{P}_C$. Denote $\sqrt{\frac{\log h_{1,n}^{-1}}{nh_{1,n}}} + h_{1,n}^2 := U(h_{1,n})$. Then we have,

$$\sup_{t\in\mathbb{R}} |\hat{f}(t; h_{1,n}) - f(t)| = \sup_{t\in\mathbb{R}} |D(t; h_{1,n}) + b(t; h_{1,n})| = O\left(U(h_{1,n})\right) \tag{10}$$

and

$$\sup_{t\in\mathbb{R}} |\delta(t)| = O\left(U(h_{1,n})\right) \tag{11}$$

for $f \in \mathcal{P}_{C,2}$. By the definition of $\delta(t)$, we also have,

$$\delta(t) = \frac{\alpha'(f(t))[\hat{f}(t;h_{1,n}) - f(t)]}{\alpha(f(t))} + \frac{\alpha''(\eta)[\hat{f}(t;h_{1,n}) - f(t)]^2}{2\alpha(f(t))} \tag{12}$$

where $\eta = \eta(t, h_{1,n}) \geq 0$ is between $\hat{f}(t;h_{1,n})$ and $f(t)$. Notice that $|\alpha''(\eta(t,h_{1,n}))| \leq c^{-3}A$ for some constant $A$ which depends only on the clipping function $p$. Set

$$L_1(t) = tK'(t) \text{ and } L(t) = K(t) + tK'(t), \quad t \in \mathbb{R}.$$

We then have the following Taylor Series expansion

$$\begin{aligned} &K\left(\frac{t-X_i}{h_{2,n}}\alpha(\hat{f}(X_i;h_{1,n}))\right) \\ &= K\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right) + L_1\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\delta(X_i) + \delta_2(t,X_i), \end{aligned} \tag{13}$$

where

$$\delta_2(t,X_i) = 1/2K''(\xi)(t-X_i)^2 h_{2,n}^{-2}\alpha^2(f(X_i))\delta^2(X_i), \tag{14}$$

$\xi$ being a (random) number between $\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))$ and $\frac{t-X_i}{h_{2,n}}\alpha(f(X_i)) + \frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\delta(X_i)$. By the analysis in Giné and Sang [4],

$$\sup_{t,x\in\mathbb{R}} |\delta_2(t,x)| = O\left(\|\hat{f}(\cdot;h_{1,n}) - f(\cdot)\|_\infty^2\right) = O(U^2(h_{1,n})) \text{ if } f \in \mathcal{P}_{C,2}. \tag{15}$$

Therefore,

$$\begin{aligned} \hat{f}(t;h_{1,n},h_{2,n}) &= \bar{f}(t;h_{2,n}) \\ &\quad + (nh_{2,n})^{-1}\sum_{i=1}^n L\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\alpha(f(X_i))\delta(X_i) \\ &\quad + (nh_{2,n})^{-1}\sum_{i=1}^n \alpha(f(X_i))\delta_2(t,X_i) \\ &\quad + (nh_{2,n})^{-1}\sum_{i=1}^n L_1\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\delta^2(X_i)\alpha(f(X_i)) \\ &\quad + (nh_{2,n})^{-1}\sum_{i=1}^n \alpha(f(X_i))\delta(X_i)\delta_2(t,X_i). \end{aligned}$$

## 3. Variance of the true estimator

To easy the notation, we denote $K_t(x) = K\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)$, $L_t(x) = L\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)$, $L_{t,1}(x) = L_1\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)$, $M(x) = L_t(x)\alpha'(f(x)) = L\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)\alpha'(f(x))$ and $M_i = M(X_i)$.

**Theorem 1** *Let $X_1, ..., X_n$ be a random sample of size n with density function $f(t)$, $t \in \mathbb{R}$, and $\hat{f}(t; h_{1,n}, h_{2,n})$ defined as in (7) is an estimator of $f(t)$. Assume that the kernel $K$ on $\mathbb{R}$ is a non-negative symmetric function with support contained in $[-T, T]$, $T < \infty$, integrates to 1 and has bounded fourth order derivatives. The function $\alpha(x)$ in the estimator $\hat{f}(t; h_{1,n}, h_{2,n})$ is defined in (3) for a nondecreasing clipping function $p(s)$ $[p(s) \geq 1$ for all s and $p(s) = s$ for all $s \geq c \geq 1]$ with five bounded and uniformly continuous derivatives, and constant $c > 0$. Then,*

$$Var(\hat{f}(t; h_{1,n}, h_{2,n})) = (V_1 + V_2 + V_3)(nh_{2,n})^{-1}(1 + o(1))$$

*where*

$$V_1 = (h_{2,n})^{-1}\mathbb{E}\left(K_t^2(X_1)\alpha^2(f(X_1))\right) = \alpha(f(t))f(t)(1 + o(1)),$$
$$V_2 = \frac{2}{h_{1,n}h_{2,n}}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))M_1 K\left(\frac{X_1 - X_2}{h_{1,n}}\right)\right],$$

*and*

$$V_3 = \frac{1}{h_{1,n}^2 h_{2,n}}\mathbb{E}\left[M_1 M_2 K\left(\frac{X_1 - X_3}{h_{1,n}}\right) K\left(\frac{X_2 - X_3}{h_{1,n}}\right)\right]. \tag{16}$$

**Proof.**

Note that,

$$Var(\hat{f}(t; h_{1,n}, h_{2,n})) = \mathbb{E}\hat{f}^2(t; h_{1,n}, h_{2,n}) - (\mathbb{E}\hat{f}(t; h_{1,n}, h_{2,n}))^2$$

of the true estimator (7), we start by calculating $\mathbb{E}\hat{f}^2(t; h_{1,n}, h_{2,n})$. It is obvious that

$$\mathbb{E}\hat{f}^2(t; h_{1,n}, h_{2,n})$$
$$= \frac{1}{nh_{2,n}^2}\mathbb{E}\left(K^2\left(\frac{t - X_1}{h_{2,n}}\alpha(\hat{f}(X_1; h_{1,n}))\right)\alpha^2(\hat{f}(X_1; h_{1,n}))\right) \tag{17}$$
$$+ \frac{n-1}{nh_{2,n}^2}\mathbb{E}\left(K\left(\frac{t - X_1}{h_{2,n}}\alpha(\hat{f}(X_1; h_{1,n}))\right)\right.$$
$$\left.\times\alpha(\hat{f}(X_1; h_{1,n}))K\left(\frac{t - X_2}{h_{2,n}}\alpha(\hat{f}(X_2; h_{1,n}))\right)\alpha(\hat{f}(X_2; h_{1,n}))\right). \tag{18}$$

The decompositions of $\alpha(\hat{f}(X_1;h_{1,n}))$ in (8) and $K\left(\frac{t-X_1}{h_{2,n}}\alpha(\hat{f}(X_1;h_{1,n}))\right)$ in (13) then give:

$$
\begin{aligned}
&(17) = (nh_{2,n}^2)^{-1}\mathbb{E}\left(K_t^2(X_1)\alpha^2(f(X_1))(1+\delta(X_1))^2\right) && (19)\\
&+ (nh_{2,n}^2)^{-1}\mathbb{E}\left(L_{t,1}^2(X_1)\alpha^2(f(X_1))\delta^2(X_1)(1+\delta(X_1))^2\right) && (20)\\
&+ (nh_{2,n}^2)^{-1}\mathbb{E}\left(\delta_2^2(t,X_1)\alpha^2(f(X_1))(1+\delta(X_1))^2\right) && \\
&+ 2(nh_{2,n}^2)^{-1}\mathbb{E}\left(K_t(X_1)\alpha^2(f(X_1))(1+\delta(X_1))^2 L_{t,1}(X_1)\delta(X_1)\right) && (21)\\
&+ 2(nh_{2,n}^2)^{-1}\mathbb{E}\left(L_{t,1}(X_1)\delta(X_1)\delta_2(t,X_1)\alpha^2(f(X_1))(1+\delta(X_1))^2\right) && \\
&+ 2(nh_{2,n}^2)^{-1}\mathbb{E}\left(K_t(X_1)\delta_2(t,X_1)\alpha^2(f(X_1))(1+\delta(X_1))^2\right). && (22)
\end{aligned}
$$

By Proposition 1 of Nakarmi and Sang [9], we have $(19) = \frac{V_1(1+o(1))}{nh_{2,n}} = \frac{\alpha(f(t))f(t)}{nh_{2,n}}(1+o(1))$. Notice that each term in (20)-(22) has $\delta(\cdot)$ or $\delta_2(\cdot)$. Then, they all have order $o(n^{-1})$ if we take $h_{1,n} = O(n^{-1/5})$ and $h_{2,n} = O(n^{-1/9})$ by the boundedness of $K$, $L_1$, $\alpha(f)$ (due to the boundedness of $f$) and (11) and (15). For example,

$$(21) = O\left(\frac{U(h_{1,n})}{nh_{2,n}^2}\right) = o(n^{-1}).$$

Again, by applying the decompositions of $\alpha(\hat{f}(X_1;h_{1,n}))$ in (8) and $K\left(\frac{t-X_i}{h_{2,n}}\alpha(\hat{f}(X_i;h_{1,n}))\right)$ in (13), and noting that the expectation terms with coefficients $-\frac{1}{nh_{2,n}^2}$ and the terms with coefficients $\frac{n}{nh_{2,n}^2} = \frac{1}{h_{2,n}^2}$ along with quantities $\delta^2(\cdot)\delta(\cdot)$, $\delta_2(\cdot)\delta(\cdot)$ or $\delta_2(\cdot)\delta_2(\cdot)$ are negligible (comparing with $O(nh_{2,n})^{-1}$) by the boundedness of $K$, $L_1$, $\alpha(f)$ and (11) and (15), we get

$$
\begin{aligned}
&(18) = (n-1)(nh_{2,n}^2)^{-1}\mathbb{E}\left[K_t(X_1)K_t(X_2)\alpha(f(X_1))\alpha(f(X_2))\right] && (23)\\
&+ h_{2,n}^{-2}\mathbb{E}\left[L_t(X_1)L_t(X_2)\delta(X_1)\delta(X_2)\alpha(f(X_1))\alpha(f(X_2))\right] && (24)\\
&+ 2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_1)\alpha(f(X_1))L_t(X_2)\delta(X_2)\alpha(f(X_2))\right] && (25)\\
&+ 2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_1)\alpha(f(X_1))L_{t,1}(X_2)\delta^2(X_2)\alpha(f(X_2))\right] && (26)\\
&+ 2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_1)\alpha(f(X_1))\delta_2(t,X_2)\alpha(f(X_2))\right] + o((nh_{2,n})^{-1}). && (27)
\end{aligned}
$$

On the other hand, by the boundedness of $K$, $L_1$, $\alpha(f)$ and (11) and (15), the terms of the form $\frac{1}{h_{2,n}}\mathbb{E}(\cdot)\mathbb{E}(\cdot)$ have order $o(\frac{1}{nh_{2,n}})$ if they include quantities $\delta^2(\cdot)\delta(\cdot)$, $\delta_2(\cdot)\delta(\cdot)$ or $\delta_2(\cdot)\delta_2(\cdot)$. Therefore,

$$
\begin{aligned}
&\left(\mathbb{E}(\hat{f}(t;h_{1,n},h_{2,n}))\right)^2 = h_{2,n}^{-2}\left\{\mathbb{E}\left[K_t(X_1)\alpha(f(X_1))\right]\right\}^2 && (28)\\
&+ h_{2,n}^{-2}\left[\mathbb{E}\left(L_t(X_1)\delta(X_1)\alpha(f(X_1))\right)\right]^2 && (29)\\
&+ 2h_{2,n}^{-2}\mathbb{E}[K_t(X_1)\alpha(f(X_1)]\mathbb{E}[L_t(X_1)\delta(X_1)\alpha(f(X_1))] && (30)\\
&+ 2h_{2,n}^{-2}\mathbb{E}[K_t(X_1)\alpha(f(X_1)]\mathbb{E}[L_{t,1}(X_1)\delta^2(X_1)\alpha(f(X_1))] && (31)\\
&+ 2h_{2,n}^{-2}\mathbb{E}[K_t(X_1)\alpha(f(X_1))]\mathbb{E}[\delta_2(t,X_1)\alpha(f(X_1))] + o(nh_{2,n})^{-1}. && (32)
\end{aligned}
$$

By the above analysis, we shall study the difference between the terms (19), (23)-(27) and (28)-(32).

**The difference between (23) and (28)**

The difference between (23) and (28) is $n^{-1}\left[\mathbb{E}\left((h_{2,n})^{-1}K_t(X_1)\alpha(f(X_1))\right)\right]^2$, which has order $O(n^{-1})$ by the bias formula of the ideal estimator.

**The difference between (24) and (29)**

Since $L$ has bounded support and $\alpha(f(X))$ is is bounded and bounded away from zero,

$$\mathbb{E}|M_1| \le \int L\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)|\alpha'(f(x))|f(x)dx = O(h_{2,n}). \tag{33}$$

Recall the decomposition of $\delta(t)$ in (12). (9), (10) and (33) give

$$h_{2,n}^{-1}\mathbb{E}\left(M_1 b(X_1;h_{1,n})\right) = O(h_{1,n}^2),$$

$$h_{2,n}^{-1}\mathbb{E}\left(M_1[\hat{f}(X_1;h_{1,n}) - f(t)]\right) = O(U(h_{1,n}))$$

and

$$h_{2,n}^{-1}\mathbb{E}\left(M_1\alpha''(\eta)[\hat{f}(X_1;h_{1,n}) - f(t)]^2\right) = O(U^2(h_{1,n})).$$

We first apply the decomposition (12) of $\delta(t)$, and use the above analysis. Then, by applying the results from section 3.1 and 3.2 of Giné and Sang [4], we get the following,

$$\begin{aligned}(29) &= \left[h_{2,n}^{-1}\mathbb{E}\left(M_1 D(X_1;h_{1,n})\right) + h_{2,n}^{-1}\mathbb{E}\left(M_1 b(X_1;h_{1,n})\right)\right]^2 + O(U^3(h_{1,n}))\\ &= \left[h_{2,n}^{-1}\mathbb{E}\left(M_1 b(X_1;h_{1,n})\right)\right]^2 + O\left(U^3(h_{1,n}) + h_{1,n}n^{-1} + n^{-2}h_{1,n}^{-2}\right). \end{aligned} \tag{34}$$

On the other hand, by a similar analysis,

$$\begin{aligned}&(24) = h_{2,n}^{-2}\mathbb{E}\left(M_1M_2[D(X_1;h_{1,n}) + b(X_1;h_{1,n})][D(X_2;h_{1,n}) + b(X_2;h_{1,n})]\right) + O(U^3(h_{1,n}))\\ &:= Q_1 + Q_2 + Q_3 + O(U^3(h_{1,n})),\end{aligned}$$

where

$$\begin{aligned}Q_1 &= h_{2,n}^{-2}\mathbb{E}(M_1M_2D(X_1;h_{1,n})D(X_2;h_{1,n}))\\ Q_2 &= h_{2,n}^{-2}\mathbb{E}(M_1M_2D(X_1;h_{1,n})b(X_2;h_{1,n}))\\ Q_3 &= h_{2,n}^{-2}\mathbb{E}(M_1M_2b(X_1;h_{1,n})b(X_2;h_{1,n})) = h_{2,n}^{-2}[\mathbb{E}(M_1b(X_1;h_{1,n}))]^2.\end{aligned}$$

We also denote $N_i = \int K\left(\frac{X_i-u}{h_{1,n}}\right) f(u)du, i = 1, 2$. Then

$$Q_1 = \frac{1}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 (K(0) - N_1)(K(0) - N_2)\right] \quad (35)$$

$$+\frac{1}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 \left(K\left(\frac{X_1 - X_2}{h_{1,n}}\right) - N_1\right)\left(K\left(\frac{X_2 - X_1}{h_{1,n}}\right) - N_2\right)\right] \quad (36)$$

$$+\frac{2(n-2)}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 \left(K\left(\frac{X_1 - X_2}{h_{1,n}}\right) - N_1\right)\left(K\left(\frac{X_2 - X_3}{h_{1,n}}\right) - N_2\right)\right] \quad (37)$$

$$+\frac{n-2}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 \left(K\left(\frac{X_1 - X_3}{h_{1,n}}\right) - N_1\right)\left(K\left(\frac{X_2 - X_3}{h_{1,n}}\right) - N_2\right)\right] \quad (38)$$

$$+\frac{(n-2)(n-3)}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 \left(K\left(\frac{X_1 - X_3}{h_{1,n}}\right) - N_1\right)\left(K\left(\frac{X_2 - X_4}{h_{1,n}}\right) - N_2\right)\right] \quad (39)$$

$$+\frac{2}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 \left(K(0) - N_1\right)\left(K\left(\frac{X_2 - X_1}{h_{1,n}}\right) - N_2\right)\right] \quad (40)$$

$$+\frac{2(n-2)}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 \left(K(0) - N_1\right)\left(K\left(\frac{X_2 - X_3}{h_{1,n}}\right) - N_2\right)\right]. \quad (41)$$

It is easy to see that, (35), (36), (40)$= O(n^{-2}h_{1,n}^{-2})$ by applying (33).
Since

$$\mathbb{E}\left(K\left(\frac{X_2 - X_3}{h_{1,n}}\right)|X_2\right) = \int K\left(\frac{X_2 - u}{h_{1,n}}\right) f(u)du = N_2, \quad (42)$$

(37)$= 0$. Similarly, $(39) = (41) = 0$.

From (3.37) of Giné and Sang [4], we have $V_3 = O(1)$. Also it is easy to see that $\mathbb{E}(M_1 N_1) = O(h_{1,n}h_{2,n})$. Hence

$$\begin{aligned}(38) =& \frac{n-2}{n^2 h_{1,n}^2 h_{2,n}^2}\mathbb{E}\left[M_1 M_2 K\left(\frac{X_1 - X_3}{h_{1,n}}\right) K\left(\frac{X_2 - X_3}{h_{1,n}}\right)\right] \\ &- \frac{n-2}{n^2 h_{1,n}^2 h_{2,n}^2}[\mathbb{E}(M_1 N_1)]^2 = \frac{V_3}{n h_{2,n}} + O(n^{-1})\end{aligned}$$

and $Q_1 = \frac{V_3}{nh_{2,n}} + O(n^{-1})$.

$$Q_2 = \frac{2}{h_{2,n}^2}\mathbb{E}\left(M_1M_2D(X_1;h_{1,n})b(X_2;h_{1,n})\right)$$

$$= \frac{2}{nh_{1,n}^2h_{2,n}^2}\mathbb{E}\left(M_1M_2\left[K\left(\frac{X_1-X_2}{h_{1,n}}\right)-N_1\right]\left[\int K\left(\frac{X_2-v}{h_{1,n}}\right)(f(v)-f(X_2))dv\right]\right) \tag{43}$$

$$+ \frac{2(n-2)}{nh_{1,n}^2h_{2,n}^2}\mathbb{E}\left(M_1M_2\left[K\left(\frac{X_1-X_3}{h_{1,n}}\right)-N_1\right]\left[\int K\left(\frac{X_2-v}{h_{1,n}}\right)(f(v)-f(X_2))dv\right]\right) \tag{44}$$

$$+ \frac{2}{nh_{1,n}^2h_{2,n}^2}\mathbb{E}\left(M_1M_2\left[K(0)-N_1\right]\left[\int K\left(\frac{X_2-v}{h_{1,n}}\right)(f(v)-f(X_2))dv\right]\right). \tag{45}$$

Under the condition that $f(x)$ has bounded second order continuous derivative,

$$\int K\left(\frac{X_2-v}{h_{1,n}}\right)(f(v)-f(X_2))dv = O(h_{1,n}^3).$$

Thus, by the boundedness of $K$ and $N_1$ and (33),

$$(43) = O(h_{1,n}n^{-1}) = (45).$$

$(44) = 0$ since

$$\mathbb{E}\left(K\left(\frac{X_1-X_3}{h_{1,n}}\right)|X_1\right) = \int K\left(\frac{X_2-u}{h_{1,n}}\right)f(u)du = N_1.$$

Therefore, $Q_2 = O(h_{1,n}n^{-1})$. Notice that the first quantity in (34) is same as $Q_3$. Hence

$$(24) - (29) = V_3(nh_{2,n})^{-1} + O(n^{-1}).$$

**The difference between (26) and (31)**

Next we denote $A(x) = K_t(x)\alpha(f(x))$ and $L_{t,2}(x) = L_{t,1}(x)\frac{\alpha'^2(f(x))}{\alpha(f(x))}$. Then (10) and the decomposition of $\delta(t)$ in (12) give

$$(26) = 2h_{2,n}^{-2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)(D(X_2;h_{1,n})+b(X_2;h_{1,n}))^2\right] + O(U^3(h_{1,n}))$$

$$= 2h_{2,n}^{-2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)D^2(X_2;h_{1,n})\right] \tag{46}$$

$$+ 4h_{2,n}^{-2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)D(X_2;h_{1,n})b(X_2;h_{1,n})\right] \tag{47}$$

$$+ 2h_{2,n}^{-2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)b^2(X_2;h_{1,n})\right] + O(U^3(h_{1,n})) \tag{48}$$

$$(46) = \frac{2}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\sum_{i=1}^n\left[K\left(\frac{X_2-X_i}{h_{1,n}}\right)-N_2\right]^2\right] \tag{49}$$

$$+ \frac{4}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\sum_{1\le i<j\le n}\left[K\left(\frac{X_2-X_i}{h_{1,n}}\right)-N_2\right]\left[K\left(\frac{X_2-X_j}{h_{1,n}}\right)-N_2\right]\right] \tag{50}$$

By independence and a similar argument as in (33), we have

$$\begin{aligned}(49) =&\frac{2}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\left[K(0)-N_2\right]^2\right]\\&+\frac{2}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\left[K\left(\frac{X_2-X_1}{h_{1,n}}\right)-N_2\right]^2\right]\\&+\frac{2(n-2)}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\left[K\left(\frac{X_2-X_3}{h_{1,n}}\right)-N_2\right]^2\right] \qquad (51)\\=&O(nh_{1,n})^{-2}+(51)\end{aligned}$$

and

$$\begin{aligned}(50) =&\frac{4}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\left[K\left(\frac{X_2-X_1}{h_{1,n}}\right)-N_2\right]\left[K(0)-N_2\right]\right]\\&+\frac{4(n-2)}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\left[K\left(\frac{X_2-X_1}{h_{1,n}}\right)-N_2\right]\left[K\left(\frac{X_2-X_3}{h_{1,n}}\right)-N_2\right]\right] \qquad (52)\\&+\frac{4(n-2)}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}\left[A(X_1)L_{t,2}(X_2)\left[K(0)-N_2\right]\left[K\left(\frac{X_2-X_3}{h_{1,n}}\right)-N_2\right]\right] \qquad (53)\\=&O(nh_{1,n})^{-2}.\end{aligned}$$

The last two terms (52) and (53) are zeroes by the argument (42). Now we decompose the term (31). Again, by (10) and the decomposition of $\delta(t)$ in (12),

$$\begin{aligned}(31) =&2h_{2,n}^{-2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)(D(X_2;h_{1,n})+b(X_2;h_{1,n}))^2\right]+O(U^3(h_{1,n}))\\=&2h_{2,n}^{-2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)D^2(X_2;h_{1,n})\right] \qquad (54)\\&+4h_{2,n}^{-2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)(D(X_2;h_{1,n})b(X_2;h_{1,n}))\right] \qquad (55)\\&+2h_{2,n}^{-2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)b^2(X_2;h_{1,n})\right]+O(U^3(h_{1,n})). \qquad (56)\end{aligned}$$

$$(54) = \frac{2}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}A(X_1)\mathbb{E}\left[L_{t,2}(X_2)\sum_{i=1}^{n}\left[K\left(\frac{X_2-X_i}{h_{1,n}}\right)-N_2\right]^2\right] \qquad (57)$$

$$+\frac{4}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}A(X_1)\mathbb{E}\left[L_{t,2}(X_2)\sum_{1\le i<j\le n}\left[K\left(\frac{X_2-X_i}{h_{1,n}}\right)-N_2\right]\left[K\left(\frac{X_2-X_j}{h_{1,n}}\right)-N_2\right]\right] \qquad (58)$$

By a similar argument as in (33), we have

$$\begin{aligned}
(57) =&\frac{2}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)\left[K(0)-N_2\right]^2\right]\\
&+\frac{2}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)\left[K\left(\frac{X_2-X_1}{h_{1,n}}\right)-N_2\right]^2\right]\\
&+\frac{2(n-2)}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)\left[K\left(\frac{X_2-X_3}{h_{1,n}}\right)-N_2\right]^2\right] \qquad (59)\\
=&O(n^{-2}h_{1,n}^{-2})+(59)
\end{aligned}$$

and

$$\begin{aligned}
(58) =&\frac{4}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)\left[K\left(\frac{X_2-X_1}{h_{1,n}}\right)-N_2\right]\left[K(0)-N_2\right]\right]\\
&+\frac{4(n-2)}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)\left[K\left(\frac{X_2-X_1}{h_{1,n}}\right)-N_2\right]\left[K\left(\frac{X_2-X_3}{h_{1,n}}\right)-N_2\right]\right] \qquad (60)\\
&+\frac{4(n-2)}{n^2h_{1,n}^2h_{2,n}^2}\mathbb{E}[A(X_1)]\mathbb{E}\left[L_{t,2}(X_2)\left[K(0)-N_2\right]\left[K\left(\frac{X_2-X_3}{h_{1,n}}\right)-N_2\right]\right] \qquad (61)\\
=&O(n^{-2}h_{1,n}^{-2}).
\end{aligned}$$

The last two terms (60) and (61) are zeroes by the argument (42). By similar explanation as $Q_2$, $(55) = O(h_{1,n}n^{-1}) = (47)$. Also notice that (48)=(56) and (51)=(59). Therefore, we get,

$$(26) - (31) = o\left(n^{-1}\right).$$

**The difference between (27) and (32)**

To study the difference between (27) and (32), by Proposition 2 of Giné and Sang [4], for the ideal estimator (2), we have

$$||\bar{f}(t;h_{2,n}) - \mathbb{E}(\bar{f}(t;h_{2,n}))||_\infty = \sqrt{\frac{\log n}{nh_{2,n}}}.$$

Therefore,

$$||K_t(X_2)\alpha(f(X_2)) - \mathbb{E}\left(K_t(X_2)\alpha(X_2)\right)||_\infty = h_{2,n}\sqrt{\frac{\log n}{nh_{2,n}}}.$$

Hence, we have the following for the difference between (27) and (32).

$$\begin{aligned}
(27)-(32) =& 2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\alpha(f(X_1))\delta_2(t,X_1)\right] \\
&-2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right]\mathbb{E}\left[\alpha(f(X_1))\delta_2(t,X_1)\right] \\
=&2h_{2,n}^{-2}\mathbb{E}\{\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\alpha(f(X_1))\delta_2(t,X_1)|X_2\right]\} \\
&-2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right]\mathbb{E}\{\mathbb{E}\left[\alpha(f(X_1))\delta_2(t,X_1)|X_2\right]\} \\
=&2h_{2,n}^{-2}\mathbb{E}\{(K_t(X_2)\alpha(f(X_2))-\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right])\mathbb{E}\left[\alpha(f(X_1))\delta_2(t,X_1)|X_2\right]\} \\
\le&\frac{2}{h_{2,n}}\sqrt{\frac{\log n}{nh_{2,n}}}\sup_{t,x\in\mathbb{R}}|\delta_2(t,x)| \\
=&\sqrt{\frac{\log n}{nh_{2,n}^3}}\;O\left(\left(h_{1,n}^2+\sqrt{\frac{\log n}{nh_{1,n}}}\right)^2\right)=o(n^{-1}).
\end{aligned}$$

**The difference between (25) and (30)**

We shall apply the decomposition of $\delta(t)$ as in (12). The difference between $(25)$ and $(30)$ with the second part of (12) has order $o(n^{-1})$ as the analysis in the difference between (27) and (32). Therefore

$$\begin{aligned}
(25)-(30) =&2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))M_1(D(X_1;h_{1,n})+b(X_1;h_{1,n}))\right] \\
&-2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(X_2)\right]\mathbb{E}\left[M_1(D(X_1;h_{1,n})+b(X_1;h_{1,n}))\right]+o(n^{-1}) \\
=&2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))M_1D(X_1;h_{1,n})\right] \\
&-2h_{2,n}^{-2}\mathbb{E}\left[K_t(X_2)\alpha(X_2)\right]\mathbb{E}\left[M_1D(X_1;h_{1,n})\right]+o(n^{-1}) \\
=&\frac{2}{nh_{1,n}h_{2,n}^2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))M_1\left(K\left(\frac{X_1-X_2}{h_{1,n}}\right)-N_1\right)\right] \qquad (62)\\
&-\frac{2}{nh_{1,n}h_{2,n}^2}\mathbb{E}\left[K_t(X_2)\alpha(X_2)\right]\mathbb{E}\left[M_1\left(K\left(\frac{X_1-X_2}{h_{1,n}}\right)-N_1\right)\right]+o(n^{-1}) \qquad (63)\\
=&\frac{2}{nh_{1,n}h_{2,n}^2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))M_1K\left(\frac{X_1-X_2}{h_{1,n}}\right)\right] \qquad (64)\\
&-\frac{2}{nh_{1,n}h_{2,n}^2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right]\mathbb{E}(M_1N_1)+o(n^{-1}). \qquad (65)
\end{aligned}$$

We have the equality (62) by the independence. The term (63) is zero.

Since $K, L\alpha'$ and $f$ are bounded functions and $K(v)$ has bounded support, we have

$$\begin{aligned}
|V_2| =& \frac{2}{h_{1,n}h_{2,n}}\left|\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\int L_t(u)\alpha'(f(u))K\left(\frac{u-X_2}{h_{1,n}}\right)f(u)du\right]\right| \\
=& \frac{2}{h_{2,n}}\left|\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\int L_t(X_2+h_{1,n}v)\alpha'(f(X_2+h_{1,n}v))K(v)f(X_2+h_{1,n}v)dv\right]\right| \\
\le& \frac{C}{h_{2,n}}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right]=O\left(1\right) \qquad (66)
\end{aligned}$$

and $(64) = O(nh_{2,n})^{-1}$. On the other hand,

$$\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right] = O(h_{2,n})$$

and

$$\begin{aligned}\mathbb{E}(M_1N_1) =& \int L_t(u)\alpha'(f(u))f(u)\int K\left(\frac{u-v}{h_{1,n}}\right)f(v)dvdu\\ =& h_{1,n}\int L_t(u)\alpha'(f(u))f(u)\int K(w)f(u-h_{1,n}w)dwdu\\ =& O(h_{1,n}h_{2,n})\end{aligned}$$

since the function $\alpha(f)$ in $L_t$ is bounded below and above. Therefore in (65),

$$\frac{2}{nh_{1,n}h_{2,n}^2}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\right]\mathbb{E}(M_1N_1) = O(n^{-1}). \tag{67}$$

Hence (66) and (67) give us

$$(25) - (30) = V_2(nh_{2,n})^{-1} + O(n^{-1}).$$

Thus,

$$Var(\hat{f}(t; h_{1,n}, h_{2,n})) = (V_1 + V_2 + V_3)(nh_{2,n})^{-1}(1 + o(1)).$$

■

## 4. Optimal bandwidth selection

Denote $z_n = \left(w - \frac{h_{1,n}v}{h_{2,n}}\right)\alpha(f(h_{1,n}v + t - h_{2,n}w))$. Then $\lim_{n\to\infty} z_n = w\alpha(f(t))$ since $h_{1,n}/h_{2,n} \to 0$. In the following we change variables twice, then take limit by applying the compact support of $K$ and the boundedness below and above of the function $\alpha$.

$$\begin{aligned}V_2 =& \frac{2}{h_{1,n}h_{2,n}}\mathbb{E}\left[K_t(X_2)\alpha(f(X_2))\int M(u)K\left(\frac{u-X_2}{h_{1,n}}\right)f(u)du\right]\\ =& 2\int K\left(w\alpha(f(t-h_{2,n}w))\right)\alpha(f(t-h_{2,n}w))f(t-h_{2,n}w)\\ &\times \int_{-T}^{T} L(z_n)\alpha'(f(h_{1,n}v+t-h_{2,n}w))K(v)f(h_{1,n}v+t-h_{2,n}w)dvdw\\ \xrightarrow{n\to\infty}& 2\int K(w\alpha(f(t)))\alpha(f(t))f(t)\int_{-T}^{T} L\left(w\alpha(f(t))\right)\\ &\times \alpha'(f(t))K(v)f(t)dvdw\\ =& 2\alpha'(f(t))f^2(t)\int K(y)L(y)dy\\ =& \alpha'(f(t))f^2(t)\mu_0 := V_2'.\end{aligned}$$

For the quantity $V_3$, let $x_3 = x_1 - h_{1,n}s$, $x_1 = x_2 - h_{1,n}w$ and $x_2 = t - h_{2,n}z$ be the change of variables. Again, since $h_{1,n}/h_{2,n} \to 0$, $L$ has compact support and the function $\alpha$ is bounded below and above,

$$\begin{aligned}
(16) &= \frac{1}{h_{1,n}^2 h_{2,n}} \int\int\int L_t(x_1)\alpha'(f(x_1))f(x_1)L_t(x_2)\alpha'(f(x_2))f(x_2) \\
&\times K\left(\frac{x_1 - x_3}{h_{1,n}}\right) K\left(\frac{x_2 - x_3}{h_{1,n}}\right) f(x_3)dx_3dx_1dx_2 \\
&= \int\int\int L\left(\left(z + \frac{wh_{1,n}}{h_{2,n}}\right)\alpha(f(t - h_{2,n}z - h_{1,n}w))\right)\alpha'(f(t - h_{2,n}z - h_{1,n}w)) \\
&\times f(t - h_{2,n}z - h_{1,n}w)L(z\alpha(f(t - h_{2,n}z))K(w + s)\alpha'(f(t - h_{2,n}z))f(t - h_{2,n}z) \\
&\times f(t - h_{2,n}z - h_{1,n}w - h_{1,n}s)K(s)dwdzds \\
&\xrightarrow{n\to\infty} f^3(t)\alpha'^2(f(t))\int\int\int L^2(z\alpha(f(t)))K(s)K(w + s)dwdzds \\
&= f^3(t)\alpha'^2(f(t))\int L^2(z\alpha(f(t)))dz \\
&= f^3(t)\frac{\alpha'^2(f(t))}{\alpha(f(t))}\int L^2(y)dy := V_3'.
\end{aligned}$$

**Theorem 2** *Suppose $f$ is a density in $C^4(\mathbb{R})$ and $p$ is a clipping function in $C^5(\mathbb{R})$. Then the integrated mean squared error on $D_r$ is*

$$R(h_{1,n}, h_{2,n})|D_r = \int_{D_r}\left(\frac{\tau_4 D_4(1/f)}{24}\right)^2 h_{2,n}^8 dt + \int_{D_r}\frac{V_1 + V_2 + V_3}{nh_{2,n}}dt + o(h_{2,n}^8). \qquad (68)$$

*Furthermore, the optimal bandwidth $h_{2,n}$ is given by*

$$h_{*(2,n)} = \left[\frac{576(V_1 + V_2' + V_3')}{n\tau_4^2(D_4(1/f))^2}\right]^{1/9}.$$

**Proof.** The bias in the integrated MSE (68) is from Corollary 1 in Giné and Sang [4]. The variance is dominated by $\frac{V_1+V_2+V_3}{nh_{2,n}}$. They are from the analysis of (19), the differences between (23)-(27) and (28)-(32) respectively. ■

## References


[1] Abramson, I. 1982. On bandwidth variation in kernel estimates - a square-root law, *Ann. Statist.* **10** 1217-1223.

[2] Giné, E. and A. Guillou. 2002. Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. I. H. Poincaré-Probab. et Stat.* **38** 907-921.

[3] Giné, E. and H. Sang. 2010. Uniform asymptotics for kernel density estimators with variable bandwidths. *J. Nonparametric Statistics* **22** 773-795.

[4] Giné, E. and H. Sang. 2013. On the estimation of smooth densities by strict probability densities at optimal rates in sup-norm. *IMS Collections, From Probability to Statistics and Back: High-Dimensional Models and Processes* **9** 128-149.

[5] Hall, P. and J. S. Marron. 1988. Variable window width kernel estimates of probability densities. *Probab. Th. Rel. Fields* **80** 37-49. Erratum: *Probab. Th. Rel. Fields* **91** 133.

[6] P. Hall, T. Hu and J. S. Marron. Improved Variable Window Kernel Estimates of Probability Densities. *Ann. Statist.* **23** (1995) 1-10.

[7] McKay, I. J. 1993. A note on bias reduction in variable kernel density estimates. *Canad. J. Statist.* **21** 367-375.

[8] McKay, I.J.1 993.Variable kernel methods in density estimation. Ph.D Dissertation, Queen's University.

[9] Nakarmi, J. and H. Sang. 2014. Central limit theorem for the variable bandwidth kernel estimator. Submitted.

[10] S. Yu. Novak. 1999.A generalized kernel density estimator. (Russian) *Teor. Veroyatnost. i Primenen.* 2000.**44** 634–645; translation in Theory Probab. Appl. **44** 570–583.

[11] Terrell, G. R. and D. Scott. 1992.Variable kernel density estimation. *Ann. Statist.* **20** 1236-1265.