Central limit theorem for the variable bandwidth kernel density estimators

Janet Nakarmi\textsuperscript{a} and Hailin Sang\textsuperscript{b}\textsuperscript{1}

\textsuperscript{a} Department of Mathematics, University of Central Arkansas, Conway, AR 72035, USA. E-mail address: janetn@uca.edu

\textsuperscript{b} Department of Mathematics, The University of Mississippi, University, MS 38677, USA. E-mail address: sang@olemiss.edu

Abstract

In this paper we study the ideal variable bandwidth kernel density estimator introduced by McKay [9, 10] and Jones et al. [7] and the plug-in practical version of the variable bandwidth kernel estimator with two sequences of bandwidths as in Giné and Sang [4]. Based on the bias and variance analysis of the ideal and plug-in variable bandwidth kernel density estimators, we study the central limit theorems for each of them. The simulation study confirms the central limit theorem and demonstrates the advantage of the plug-in variable bandwidth kernel method over the classical kernel method.

MSC 2010 subject classification: 62G07, 62E20, 62H12, 60F05

Key words and phrases: central limit theorem, variable bandwidth kernel density estimation.

1 Introduction

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

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

where $h_n$ is the bandwidth sequence with $h_n \to 0$, $nh_n^d \to \infty$, and its properties have been well studied in the literature. The variance of (1) has order $O((nh_n^d)^{-1})$ and the bias has order $O(h_n^2)$ if $f(t)$ has bounded second order partial derivatives. See Silverman [12] and Wand and Jones [14] for the literature on kernel density estimation. For $k = \ldots$
Multi-dimensional version of the variable bandwidth kernel density estimator proposed by Marron [8]. In this paper we study the following multi-dimensional version of the variable bandwidth kernel density estimator proposed by McKay [9, 10]:

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

(2)

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

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

(3)

The function $p$ has at least fourth order derivative 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 a fixed number $c$, where $0 < c < \infty$. The equation (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]. Abramson proposed the following estimator

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

(4)

where $\gamma(t, s) = (f(s) \setminus f(t)/10)^{1/2}$. 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 $f(X_i) \geq t_0c^2$ by the definition of the function $p(x)$. The estimator (2) or (4) has clipping procedure in (3) 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 procedures 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 and the clipping procedure improve 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 keep the variance at the order of $(nh_n^d)^{-1}$ if $f(t) \neq 0$ and $f(x)$ has fourth order continuous derivatives at $t$. So, one has a non-negative estimator of the density that performs asymptotically as a kernel estimator based on a fourth order (hence, partly negative) kernel. 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.

Terrell and Scott [13] and McKay [10] showed that the following modification of the Abramson estimator without the ‘clipping filter’ $(f(t)/10)^{1/2}$ on $f^{1/2}(X_i)$ studied in Hall and Marron [5],

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

(5)
which has integral 1 and thus 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. In the case $d = 1$, Hall, Hu and Marron [6] proposed the estimator

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

(6)

where $B$ is a fixed constant; see also Novak [11] 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. However, notice that this estimator and all the other variable bandwidth kernel density estimators are not applicable 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

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

by plugging in a pilot estimator, the classical estimator (1), into the estimator (5). 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), i.e.,

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

They took the Taylor expansion of $K \left( \frac{t - X_i}{h_{2,n}} \right)$ at $K \left( \frac{t - X_i}{h_{2,n}} \right)$ and then proved that the discrepancy between the plug-in estimator $\hat{f}_{HM}(t; h_{1,n}, h_{2,n})$ and the ideal version (5) has asymptotic convergence rate $O_P(n^{-4/(8+d)})$ point-wise. By applying this Taylor decomposition, McKay [10] studied convergence of plug-in estimator of (2) in probability and point-wise. Giné and Sang [3, 4] studied plug-in estimators of (6) and (2) for one and $d$-dimensional observations. They proved that the discrepancy between the plug-in estimator and the true value converges uniformly over a data adaptive region at a rate of $O_{a.s.}(\sqrt{(\log n/n)^{4/(8+d)}})$ by applying empirical process techniques. The plug-in estimator in Giné and Sang [4] has the following form

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

(7)

In this paper, we concentrate on the study of central limit theorem of the plug-in estimator (7).

The paper has the following structure. Section 2 introduces the decompositions which will be applied throughout the paper. Section 3 gives the exact bias formula. In
Section 4, we obtain an exact formula for the variance of the ideal estimator. Based on the study in Sections 3 and 4, we provide a central limit theorem for the plug-in estimator in Section 5. The simulation study in Section 6 confirms the central limit theorem and demonstrates the advantage of the plug-in variable bandwidth kernel method over the classical kernel method.

2 Preliminary decomposition

For convenience, we adopt the notations as in Giné and Sang [4] for the Taylor series expansion of \( K \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i); h_{1,n}) \right) \) at \( K \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i)) \right) \). We also give statements without detailed explanation. For details, readers are referred to Giné and Sang [4]. \( \mathcal{P}_{C,k} \) will denote the set of densities on \( \mathbb{R}^d \) for which they and their partial derivatives of order \( k \) or lower are bounded by \( C < \infty \) and are uniformly continuous. We say that a function \( g \) is in \( C_l(\Omega) \) if it and its first \( l \) derivatives are bounded and uniformly continuous on \( \Omega \).

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

\[
|\delta(t)| \leq B c^{-2} |\hat{f}(t; h_{1,n}) - f(t)| \tag{9}
\]

for a constant \( B \) that depends only on the function \( p \). Here the constant \( c \) and the function \( p \) are applied in the definition of \( \alpha(\cdot) \) in (3). Although we study the asymptotics of the true estimator point-wise, the uniform asymptotic behavior of the quantity \( \delta(\cdot) \) is needed in the latter analysis. Define

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

Note that for \( f \in \mathcal{P}_{C,2}, \sup_{t \in \mathbb{R}^d} |b(t; h_{1,n})| = O(h_{1,n}^2) \), and by Giné and Guillou [2],

\[
\sup_{t \in \mathbb{R}^d} |D(t; h_{1,n})| = O_{a.s.} \left( \sqrt{\frac{\log h_{1,n}^{-1}}{nh_{1,n}^d}} \right)
\]

for \( f \in \mathcal{P}_{C,0} \). Denote

\[
\sqrt{\frac{\log h_{1,n}^{-1}}{nh_{1,n}^d}} + h_{1,n}^2 := U(h_{1,n}). \tag{10}
\]

Then we have,

\[
\sup_{t \in \mathbb{R}^d} |\hat{f}(t; h_{1,n}) - f(t)| = \sup_{t \in \mathbb{R}^d} |D(t; h_{1,n}) + b(t; h_{1,n})| = O_{a.s.}(U(h_{1,n}))
\]
and
\[ \sup_{t \in \mathbb{R}^d} |\delta(t)| = O_{a.s.} \left( U(h_{1,n}) \right) \]  
for \( f \in \mathcal{P}_{C,2} \). By the definition of \( \delta(t) \), we also have,
\[ \delta(t) = \frac{\alpha'(f(t))[\hat{f}(t; h_{1,n}) - f(t)]}{\alpha(f(t))} + \frac{\alpha''(\eta)[\hat{f}(t; h_{1,n}) - f(t)]^2}{2\alpha(f(t))} \]  
(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 \( p \). It is also convenient to record the following expansion of \( \alpha^d(\hat{f}) \) implied by (8) and (11):
\[ \alpha^d(\hat{f}(t; h_{1,n})) = \alpha^d(f(t))(1 + d\delta(t)) + \delta_1(t) \]  
(13)
with
\[ \|\delta_1\|_\infty = O_{a.s.}(\|\delta\|^2_\infty) \quad \text{for} \quad f \in \mathcal{P}_{C,2}. \]
Hence, by (9) and (11),
\[ \|\delta_1\|_\infty = O_{a.s.}(\|\hat{f}_n(\cdot; h_{1,n}) - f(\cdot)\|^2_\infty) \quad \text{for} \quad f \in \mathcal{P}_{C,2}. \]

Set
\[ L_1(t) = \sum_{i=1}^d t_i K'_i(t) \quad \text{and} \quad L(t) = dK(t) + L_1(t), \quad t \in \mathbb{R}^d, \]  
(14)
where \( K'_i \) denotes the partial derivative of \( K \) in the direction of the \( i \)-th coordinate, and \( t_i \) denotes the \( i \)-th coordinate of \( t \in \mathbb{R}^d \). By symmetry and integration by parts, we notice that \( L \) is a second order kernel.

We then have the following Taylor series expansion
\[ K \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i; h_{1,n})) \right) = K \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i)) \right) \]
\[ + \sum_{j=1}^d K'_j \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i)) \right) \frac{(t-X_i)_j}{h_{2,n}} \alpha(f(X_i)) \delta(X_i) + \delta_2(t; X_i), \]  
(15)
where
\[ \delta_2(t, X_i) = \sum_{j, \ell=1}^d K''_{j,\ell}(\xi) \frac{(t-X_i)_j(t-X_i)_{\ell}}{2h_{2,n}^2} \alpha^2(f(X_i)) \delta^2(X_i), \]
\( \xi \) 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],
\[ \sup_{t, x \in \mathbb{R}^d} |\delta_2(t, x)| = O_{a.s.} \left( \|\hat{f}(\cdot; h_{1,n}) - f(\cdot)\|^2_\infty \right) = O_{a.s.}(U^2(h_{1,n})) \]  
(16)
if \( f \in \mathcal{P}_{C,2}. \) Therefore using equation (14), Taylor series expansion of \( K \) in (15), and expansion of \( \alpha^d \) in (13), we have

\[
\hat{f}(t; h_{1,n}, h_{2,n}) = \bar{f}(t; h_{2,n}) + \frac{1}{nh_{2,n}^d} \sum_{i=1}^{n} \left[ K \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i)) \right) \delta_1(X_i) + \alpha^d(f(X_i)) \delta_2(t, X_i) \right] + \frac{1}{nh_{2,n}^d} \sum_{i=1}^{n} \left[ L_1 \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i)) \right) \delta(X_i) \delta_1(X_i) + \alpha^d(f(X_i)) \delta(X_i) \delta_2(t, X_i) \right] + \frac{1}{nh_{2,n}^d} \sum_{i=1}^{n} \delta_2(t, X_i) \delta_1(X_i).
\]

\(3\) Bias

The following notations are necessary for the rest of the paper: for \( v = (v_1, \ldots, v_d) \in (\mathbb{N} \cup \{0\})^d \) and vector \( u = (u_1, \ldots, u_d)^T \), set

\[
|v| = \sum_{i=1}^{d} v_i, \quad v! = v_1! \cdots v_d!,
\]

\[
D_v = D_{u_1}^{v_1} \circ \cdots \circ D_{u_d}^{v_d}, \quad u^v = u_1^{v_1} \cdots u_d^{v_d},
\]

\[
\tau_v = \int_{\mathbb{R}^d} u^v K(u) du, \quad \mu_v = \int_{\mathbb{R}^d} u^v K^2(u) du,
\]

where \( D_v \) means that we take \( v_1 \) partial order derivatives on the first coordinate, \( v_2 \) partial order derivatives on the second coordinate, until we take \( v_d \) partial order derivatives on the \( d \)-th coordinate.

We also define

\[
D_r := \{ t \in \mathbb{R}^d : f(t) > r > t_0 c^2, \|t\| < 1/r \}, \quad r > 0.
\]

Here, \( c \) and \( t_0 \) are the constants that appear in the definition of the clipping function \( \alpha \) in (3).

**Proposition 3.1** Let \( f \) be a density function in \( \mathcal{P}_{C,4} \), let \( p \) be a clipping function in \( C^5(\mathbb{R}) \), set \( \alpha(f(t)) = cp^{1/2}(c^{-2}f(t)) \) for some \( c > 0 \), and define \( \hat{f}(t; h_{1,n}, h_{2,n}) \) as in (7). Suppose that the kernel \( K \) on \( \mathbb{R}^d \) has the form \( K(t) = \Phi(\|t\|^2) \) for some real function
Φ with uniformly bounded second order derivative and with support contained in $[0, T]$, $T < \infty$. $K$ is non-negative and integrates to 1. For the quantity $U(h_{1,n})$ defined in (10), assume that $U(h_{1,n}) = o(h_{2,n}^2)$. Then as $h_{2,n} \to 0$, for $t \in D_r$,

$$\mathbb{E}(\hat{f}(t; h_{1,n}, h_{2,n})) - f(t) = \left( \sum_{|v| = 4} \tau_v D_v (1/f)/v! \right) h_{2,n}^4 + o(h_{2,n}^4).$$

**Proof.** By McKay [9, 10] or Corollary 1 of Giné and Sang [4], the ideal estimator (2) satisfies

$$\mathbb{E}\hat{f}(t; h_{2,n}) = f(t) + \left( \sum_{|v| = 4} \tau_v D_v (1/f)/v! \right) h_{2,n}^4 + o(h_{2,n}^4). \quad (22)$$

Hence by the expansion of $\hat{f}(t; h_{1,n}, h_{2,n})$ in (17)-(20) and equation (22), the expectation of $\hat{f}(t; h_{1,n}, h_{2,n})$ is

$$\mathbb{E}\hat{f}(t; h_{1,n}, h_{2,n}) = f(t) + \left( \sum_{|v| = 4} \tau_v D_v (1/f)/v! \right) h_{2,n}^4 + o(h_{2,n}^4) \quad (23)$$

$$+ \frac{1}{h_{2,n}^4} \mathbb{E} \left( L \left( \frac{t - X_1}{h_{2,n}} \alpha(f(X_1)) \right) \alpha^d(f(X_1)) \delta(X_1) \right) \quad (24)$$

$$+ \mathbb{E} \left[ \frac{1}{n h_{2,n}} \sum_{i=1}^{n} \left( K \left( \frac{t - X_i}{h_{2,n}} \alpha(f(X_i)) \right) \delta_1(X_i) + \alpha^d(f(X_i)) \delta_2(t, X_i) \right) \right. \right.$$ 

$$+ dL_1 \left( \frac{t - X_i}{h_{2,n}} \alpha(f(X_i)) \right) \alpha^d(f(X_i)) \delta^2(X_i) \right] \quad (25)$$

$$+ \mathbb{E} \left[ \frac{1}{n h_{2,n}^2} \sum_{i=1}^{n} \left( L_1 \left( \frac{t - X_i}{h_{2,n}} \alpha(f(X_i)) \right) \delta(X_i) \delta_1(X_i) + d\alpha^d(f(X_i)) \delta(X_i) \delta_2(t, X_i) \right) \right. \right.$$ 

$$+ \mathbb{E} \left[ \frac{1}{n h_{2,n}^2} \sum_{i=1}^{n} \delta_2(t, X_i) \delta_1(X_i) \right]. \quad (26)$$

By (3.20) of Giné and Sang [4] and the boundedness of $\alpha$, $K$ and $L_1$, we have

$$|(25)| = O \left( U^2(h_{1,n}) \right), \quad (28)$$

$$|(26)| = O \left( U^3(h_{1,n}) \right), \quad (29)$$

and

$$|(27)| = O \left( U^4(h_{1,n}) \right), \quad (30)$$
where the $U(h_{1,n})$ is defined in (10). We further decompose (24) using the decomposition (12) of $\delta(t)$ first, and then the decomposition of $\hat{f} - f$ into the random part $D$ and the bias $b$:

\begin{equation}
(24) = \frac{1}{dh_{2,n}^d} \mathbb{E} \left[ L \left( \frac{t-X_1}{h_{2,n}} \alpha(f(X_1)) \right) \left( \alpha^d \right)'(f(X_1)) D(X_1; h_{1,n}) \right] + \frac{1}{dh_{2,n}^d} \mathbb{E} \left[ L \left( \frac{t-X_1}{h_{2,n}} \alpha(f(X_1)) \right) \left( \alpha^d \right)'(f(X_1)) b(X_1; h_{1,n}) \right] + \frac{1}{2h_{2,n}^d} \mathbb{E} \left[ \left( L \left( \frac{t-X_1}{h_{2,n}} \alpha(f(X_1)) \right) \left( \alpha^{d-1} \right)(f(X_1)) \alpha''(\eta(X_1)) [\hat{f}(X_1; h_{1,n}) - f(X_1)]^2 \right) \right].
\end{equation}

By (16) or (3.24) of Giné and Sang [4] and the boundedness of $\alpha''(\eta)$ and $L$, we obtain,

\begin{equation}
| (33) | = O \left( U^2(h_{1,n}) \right).
\end{equation}

By (3.26) of Giné and Sang [4], we have

\begin{equation}
| (32) | = O(h_{1,n}^2 h_{2,n}^2) \quad \text{for } f \in \mathcal{P}_{C;4}.
\end{equation}

In the following we give the estimation of (31) to finish the proof of the proposition.

Let $H$ be an integrable function of two i.i.d. random variables $X$ and $Y$. Then the $U$-statistic is

\[ U_n(H) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} H(X_i, X_j), \]

where the variables $X_i$ are i.i.d. copies of $X$. The second order Hoeffding projection of $H(X, Y)$ is $\pi_2(H)(X, Y) = H(X, Y) - \mathbb{E}_X H(X, Y) - \mathbb{E}_Y H(X, Y) + \mathbb{E}H$. If we set

\[ H_t(X, Y) := L \left( \frac{t-X}{h_{2,n}} \alpha(f(X)) \right) \left( \alpha^d \right)'(f(X)) K \left( \frac{X-Y}{h_{1,n}} \right), \]

then we can decompose the following quantity into a diagonal term and a $U$-statistic term,

\begin{equation}
\frac{1}{nh_{2,n}^d} \sum_{i=1}^n L \left( \frac{t-X_i}{h_{2,n}} \alpha(f(X_i)) \right) \left( \alpha^d \right)'(f(X_i)) D(X_i; h_{1,n})
\end{equation}

\begin{equation}
= \frac{1}{n^2 h_{1,n}^d h_{2,n}^d} \sum_{i=1}^n (H_t(X_i, X_i) - \mathbb{E}_X H_t(X_i, X))
\end{equation}

\begin{equation}
+ \frac{n-1}{nh_{1,n}^d h_{2,n}^d} U_n \left( \pi_2(H_t(\cdot, \cdot)) \right)
\end{equation}

\begin{equation}
+ \frac{n-1}{n^2 h_{1,n}^d h_{2,n}^d} \sum_{i=1}^n (\mathbb{E}_X H_t(X, X_i) - \mathbb{E}H_t).
\end{equation}
Obviously, (38) and (39) have mean zero since
\[ \mathbb{E} U_n (\pi_2 (H_t (\cdot , \cdot ))) = \mathbb{E} \mathbb{E}_X H_t (X, Y) - \mathbb{E} H_t = 0. \]
In the empirical process (37), set \( \bar{Q}_i (t) = H_t (X_i, X_i) - \mathbb{E}_Y H_t (X_i, Y) \) and observe that,
\[ \mathbb{E} |\bar{Q}_1 (t)| \leq B h_{2,n}^d, \]
for some finite constant \( B \). By combining the above analysis and by the analysis of Giné and Sang [4] on page 144, we have
\[ (31) = \frac{1}{d} \mathbb{E} ((36)) = \frac{1}{d} \mathbb{E} \left( \frac{1}{n^2 h_{1,n}^d h_{2,n}^d} \sum_{i=1}^n (H_t (X_i, X_i) - \mathbb{E}_Y H_t (X_i, Y)) \right) = O \left( \frac{1}{nh_{1,n}^d} \right). \]

By the analysis in (23)-(35) and (40), the bias
\[ \mathbb{E} (\hat{f} (t; h_{1,n}, h_{2,n})) - f (t) = \left( \sum_{|v| = 4} \tau_v D_v (1/f) / v! \right) h_{2,n}^4 + o(h_{2,n}^4). \]

4 Variance of the ideal estimator

We develop the second moment expansion uniformly to deal with the variance of the ideal estimator. Here we denote \( h = h_{2,n} \) and \( \gamma (s) = \alpha (f(s)) \) for convenience. Then, the ideal estimator \( (2) \) has the form
\[ \bar{f} (t) = \frac{1}{n h^d} \sum_{i=1}^n \gamma^d (X_i) K (h^{-1} \gamma (X_i) (t - X_i)), \quad t \in \mathbb{R}^d. \]
Denote \( A (X_i) = \gamma^d (X_i) K (h^{-1} \gamma (X_i) (t - X_i)) \), then we have the second moment of the ideal estimator as follows:
\[ \mathbb{E} \bar{f}^2 (t; h) = \frac{1}{n^2 h^{2d}} \mathbb{E} \left( \sum_{i=1}^n A (X_i) \right)^2 = \frac{1}{n^2 h^{2d}} \sum_{i=1}^n \mathbb{E} A^2 (X_i) + \frac{1}{n^2 h^{2d}} \sum_{i \neq j} \mathbb{E} A (X_i) \mathbb{E} A (X_j) = \frac{1}{n h^{2d}} \mathbb{E} A^2 (X_1) + \frac{n(n - 1)}{n^2 h^{2d}} (\mathbb{E} A (X_1))^2. \]
Recall that 
\[ \mathbb{E} f(t; h) = f(t) + \left( \sum_{|v| = 1} \tau_v D_v(1/f)/v! \right) h^4 + o(h^4) \]
if \( f(t) > t_0 c^2 \). Since \( \mathbb{E} A(X_1) = h^d \mathbb{E} f(t; h) \), we have

\[ \text{Var} f(t; h) = \mathbb{E} f^2(t; h) - [\mathbb{E} f(t; h)]^2 \]
\[ = \frac{1}{nh^{2d}} \mathbb{E} A^2(X_1) + \frac{(n-1)}{nh^{2d}} (\mathbb{E} A(X_1))^2 - \frac{1}{h^{2d}} (\mathbb{E} A(X_1))^2 \]
\[ = \frac{1}{nh^{2d}} \mathbb{E} A^2(X_1) - \frac{1}{n} \left[ f(t) + \left( \sum_{|v| = 1} \tau_v D_v(1/f)/v! \right) h^4 + o(h^4) \right]^2 \]
\[ = \frac{1}{nh^{2d}} \mathbb{E} A^2(X_1) + O(n^{-1}). \] (41)

In the next proposition, we study the quantity \( \mathbb{E} A^2(X_1) \). The idea is similar to the uniform bias expansion as in McKay [10], Jones, McKay and Hu [7], and particularly Giné and Sang [4].

**Proposition 4.1** Suppose that the kernel \( K \) on \( \mathbb{R}^d \) has the form \( K(t) = \Phi(||t||^2) \) for some real function \( \Phi \) with uniformly bounded second order derivative and with support contained in \( [0, T] \), \( T < \infty \). \( K \) is non-negative and integrate to 1. Assume the density function \( f \) is in \( C^d(\mathbb{R}^d) \). Suppose that \( \gamma(t) \geq c > 0 \) for some \( c > 0 \) and all \( t \in \mathbb{R}^d \), and that the function \( \gamma(t) \) is in \( C^{d+1}(\mathbb{R}^d) \). Then we have,

\[ \mathbb{E} A^2(X_1) = \sum_{k=0}^{l} a_k(t) h^{k+d} + o(h^{l+d}) \] (42)
as \( h \to 0 \), uniformly in \( t \in \mathbb{R}^d \). The set of functions \( a_k \), which are uniformly bounded and equicontinuous, are defined as

\[ a_{2k+1}(t) = 0, \quad a_{2k}(t) = \sum_{|v|=2k} \frac{\mu_v}{v!} D_v \frac{f(t)}{\gamma^{|v|}(t)} \] (43)
u for \( k \leq l/2 \), in particular, \( a_0(t) = \gamma^d f(t) \mu_0 \). Here \( |v|, v!, \mu_v \) and \( D_v \) are defined in (21).

**Proof.** Note that there exists \( \delta_1 > 0 \) such that \( a_{ii} = \gamma(t-v) + v_i \frac{\partial \gamma(t-v)}{\partial v_i}, 1 \leq i \leq d \), are bounded away from zero, \( a_{ij} = v_i \frac{\partial \gamma(t-v)}{\partial v_j}, 1 \leq i \leq d, 1 \leq j \leq d, j \neq i \), are small enough for all \( t \in \mathbb{R}^d \) and \( v \in [-\delta_1, \delta_1]^d \) for functions \( \gamma \) that are bounded away from zero and their derivatives that are bounded. Hence the matrix \( A = (a_{ij})_{i,j=1}^d \) is invertible. Thus the vector function \( v \mapsto U_i(v) := v \gamma(t-v) \) is invertible on the neighborhood \( [-\delta_1, \delta_1]^d \) of \( v = 0 \) for each \( t \in \mathbb{R}^d \). By differentiation, it is easy to see that the inverse
function, say \( V_t(u) \), is \( l + 1 \) times differentiable with continuous partial derivatives. Unless \( \| t - s \|^2 \leq h^2 T / c^2 \), \( K(h^{-1}\gamma(s)(t - s)) = 0 \). Hence, the change of variables 

\[ hz = (t - s)\gamma(t - (t - s)), \] 

that is, \( t - s = V_t(hz) \), in the following integral is valid for all \( h \) small enough

\[
\mathbb{E}A^2(X_1) = \int \gamma^{2d}(s)f(s) K^2 \left( \frac{t-s}{h} \gamma(s) \right) ds
\]

(44) 

\[
= -h^d \int \gamma^{2d}(t-V_t(hz))f(t-V_t(hz))|\det(J)|K^2(z)dz
\]

where \( J \) is the partial derivative matrix of the vector function \( V_t(hz) \) with respect to \( hz \) and \( \det(J) \) is the determinant of \( J \). If we develop the function \( \gamma^{2d}(t-V_t(hz))f(t-V_t(hz))|\det(J)| \) into powers of \( hz \) and then integrate it, noting the compactness of the domain of integration and the differentiability properties of \( f \) and \( \gamma \), we have (42).

Suppose \( \psi \) is infinitely differentiable and has bounded support. Then, changing variables \( (t = s + hu) \) from \( t \) to \( u \) in (44), developing \( \psi \), changing variables once more \( (w = u\gamma(s)) \) and integrating by parts, we obtain

\[
\int \psi(t)\mathbb{E}A^2(X_1)dt = h^d \int \int \psi(s+hu)\gamma^{2d}(s)f(s)K^2(u\gamma(s))dudu \]

(45) 

\[
= h^d \int \gamma^{2d}(s)f(s) \int \psi(s+hu)K^2(u\gamma(s))duds
\]

\[
= h^d \int \gamma^{2d}(s)f(s) \int \sum_{|v|=0}^{l} \frac{D^v\psi(s)}{k!} h^{|v|}u^vK^2(u\gamma(s))duds + o(h^{l+d})
\]

\[
= h^d \int \gamma^{d}(s)f(s) \int \sum_{|v|=0}^{l} \frac{D^v\psi(s)}{v!} h^{|v|}w^v \gamma^{|v|-d}(s)K^2(w)dwdx + o(h^{l+d})
\]

\[
= \sum_{|v|=0}^{l} \mu_v h^{|v|+d} \int f(s) \frac{D^v\psi(s)}{v!} \gamma^{|v|-d}(s)ds + o(h^{l+d})
\]

\[
= \sum_{|v|=0}^{l} (-1)^{|v|} \mu_v h^{|v|+d} v!^{-1} \int \psi(s)D_v(f(s)\gamma^{|v|-d})(s)ds + o(h^{l+d}).
\]

Here \( u^v \) is defined in (21). Notice that \( \mu_{2k+1} = 0 \) for \( k \geq 0 \). Then, (43) follows by comparing the coefficients of \( h^k \) in both expansions (42) and (45). \( \blacksquare \)

Thus, the variance of the ideal estimator is 

\[
\frac{\gamma'(t)f(t)\mu}{nh^d}(1+o(1)) = \frac{\alpha D_t(f(t))\mu}{nh^d}(1+o(1))
\]

by applying Proposition 4.1 and (41).
5 Central limit theorem

5.1 Central limit theorem for ideal estimator

The ideal estimator \( \hat{f}(t; h_{2,n}) \) in (2) can be written as a sample mean of triangular array of i.i.d. random variables, i.e., \( \hat{f}(t; h_{2,n}) = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_{n,i} \), where

\[
Y_{n,i} = \frac{1}{h_{2,n}^d} K(h_{2,n}^{-1} \alpha(f(X_i))(t - X_i)) \alpha^d(f(X_i)).
\]

Notice that \( \mathbb{E}Y_{n,i}^2 < \infty \). By (41) and Proposition 4.1,

\[
\sqrt{\text{Var}(\hat{f}(t; h_{2,n}))} = (1 + o(1)) \frac{1}{\sqrt{nh_{2,n}^d}} \alpha^d(f(t)f(t) \mu_0).
\]

Hence, by the Lindeberg’s central limit theorem for triangular array of random variables, we have the following central limit theorem for the ideal estimator for all \( t \in \mathbb{R}^d \),

\[
\sqrt{\frac{nh_{2,n}^d}{n}} [\hat{f}(t; h_{2,n}) - f(t)] \xrightarrow{D} N\left(0, \alpha^d(f(t))f(t) \mu_0\right).
\]

Since \( \mathbb{E}\hat{f}(t; h_{2,n}) - f(t) = \left(\sum_{|v|=4} \tau_v D_v(1/f)/v!\right) h_{2,n}^4 (1 + o(1)) \) for \( t \in D_r \) by McKay ([9], [10]) or Corollary 1 in Giné and Sang [4],

\[
\sqrt{\frac{nh_{2,n}^d}{n}} [\mathbb{E}\hat{f}(t; h_{2,n}) - f(t)] = c_2^{(d+8)/2} \sum_{|v|=4} \tau_v D_v(1/f)/v!(1 + o(1)),
\]

if we take \( h_{2,n} = c_2 n^{-1/(8+d)} \) for some constant \( c_2 > 0 \). Note that,

\[
\hat{f}(t; h_{2,n}) - f(t) = \hat{f}(t; h_{2,n}) - \mathbb{E}\hat{f}(t; h_{2,n}) + \mathbb{E}\hat{f}(t; h_{2,n}) - f(t).
\]

Thus, by Slutsky’s theorem, for \( t \in D_r \),

\[
\sqrt{\frac{nh_{2,n}^d}{n}} [\hat{f}(t; h_{2,n}) - f(t)] \xrightarrow{D} N\left(c_2^{(d+8)/2} \sum_{|v|=4} \tau_v D_v(1/f)/v!, \alpha^d(f(t))f(t) \mu_0\right).
\]

5.2 Central limit theorem for the plug-in variable bandwidth kernel density estimator

Based on the above central limit theorems for the ideal estimator, we have the following central limit theorems for the plug-in variable bandwidth kernel density estimator.

**Theorem 5.1** Let \( X_1, \ldots, X_n \) be a random sample of size \( n \) with density function \( f(t) \), \( t \in \mathbb{R}^d \), and \( \hat{f}(t; h_{1,n}, h_{2,n}) \) defined as in (7) is an estimator of \( f(t) \). Assume \( f(t) \) to be in
\( \mathcal{P}_{c,d} \). Suppose the kernel \( K \) on \( \mathbb{R}^d \) has the form \( K(t) = \Phi(\|t\|^2) \) for some real function \( \Phi \) with uniformly bounded second order derivative and with support contained in \([0,T] \), \( T < \infty \). \( K \) is non-negative and integrates to 1. 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 \( p(s) \) \([p(s) \geq 1 \text{ for all } s \text{ and } p(s) = s \text{ for all } s \geq c \geq 1]\) with five bounded and uniformly continuous derivatives, and constant \( c > 0 \). Let \( h_{2,n} = c_2 n^{-1/(8+d)} \) for some constants \( c_2 > 0 \) and assume that \( U(h_{1,n}) = o(h_{2,n}^2) \). Then for \( t \in \mathcal{D}_r \),

\[
\sqrt{nh_{2,n}^d}[\hat{f}(t;h_{1,n},h_{2,n}) - \mathbb{E}\hat{f}(t;h_{1,n},h_{2,n})] \xrightarrow{D} N(0,\sigma_t^2)
\]  

and

\[
\sqrt{nh_{2,n}^d}[\hat{f}(t;h_{1,n},h_{2,n}) - f(t)] \xrightarrow{D} N\left(\frac{\alpha^d(f(t))\mu_0 + f^3(t)\frac{[\alpha^d(f(t))]^2}{d\alpha^d(f(t))}}{\left(\frac{\alpha^d(f(t))\mu_0 + f^3(t)\frac{[\alpha^d(f(t))]^2}{d\alpha^d(f(t))}\int_{\mathbb{R}^d} L^2(z)dz + f^2(t)(\alpha^d)'(f(t))\mu_0, \mu_0 = \int_{\mathbb{R}^d} K^2(u)du, \text{ and } L(x) = K(x) + xK'(x)}\right)^{\frac{1}{2}}\sum_{|v|=4} \tau_v D_v(1/f)/v!}, \sigma_t^2\right).
\]  

Here, \( \sigma_t^2 = \alpha^d(f(t))\mu_0 + f^3(t)\frac{[\alpha^d(f(t))]^2}{d\alpha^d(f(t))} \int_{\mathbb{R}^d} L^2(z)dz + f^2(t)(\alpha^d)'(f(t))\mu_0, \mu_0 = \int_{\mathbb{R}^d} K^2(u)du, \text{ and } L(x) = K(x) + xK'(x) \).

**Proof.** The plug-in estimator \( \hat{f}(t;h_{1,n},h_{2,n}) \) in (7) has decomposition

\[
\hat{f}(t;h_{1,n},h_{2,n}) - \mathbb{E}\hat{f}(t;h_{1,n},h_{2,n}) = \hat{f}(t;h_{1,n},h_{2,n}) - \bar{f}(t;h_{2,n}) + \bar{f}(t;h_{2,n}) - \bar{f}(t;h_{1,n},h_{2,n}) + \mathbb{E}\bar{f}(t;h_{1,n},h_{2,n}) - \mathbb{E}\bar{f}(t;h_{1,n},h_{2,n}).
\]  

Since

\[
\sqrt{nh_{2,n}^d}[\mathbb{E}\bar{f}(t;h_{1,n},h_{2,n}) - \mathbb{E}\hat{f}(t;h_{1,n},h_{2,n})] = o(h_{2,n}\sqrt{nh_{2,n}^d}) = o(1)
\]

by the analysis in Section 3, the term (51) is negligible in the central limit theorems (47) and (48). The term (49) has decomposition as in (17) - (20). We know that (18) = \( O_{a.s.}(U^2(h_{1,n})) = o_{a.s.}(h_{2,n}^d) \), (19) = \( O_{a.s.}(U^3(h_{1,n})) = o_{a.s.}(h_{2,n}^6) \) and (20) = \( O_{a.s.}(U^4(h_{1,n})) = o_{a.s.}(h_{2,n}^8) \) by (3.20) of Giné and Sang [4]. Hence they are also negligible in the central limit theorems (47) and (48).

We can further decompose (17) into the random variation part \( D \) and the bias \( b \) by
By (3.24) of Giné and Sang [4], we have that (55) = (39) divided by other terms are all negligible. Now let Z prove the central limit theorem (47), it suffices to derive a central limit theorem for (37) = are bounded by Ch o and by the proof of (3.33) of Giné and Sang [4], (38) = d (53) multiplied by E o. Define Let ϵ>0 be given. Then, by Markov’s inequality, we have

\[ P \left( \frac{B_n}{h_{2,n}^2} > \epsilon \right) \leq \frac{\mathbb{E}(B_n^2)}{\epsilon^2 h_{2,n}^8} \leq \frac{C}{n^2 h_{1,n}^{2d} \epsilon^2 h_{2,n}^8} \xrightarrow{n \to \infty} 0. \]

Hence, (37)=o_p(h_{2,n}^4).

By the above analysis, only the term (50) and the remaining term from (49), i.e., (39) divided by d, have contribution in the central limit theorems (47) and (48). The other terms are all negligible. Now let Z_n,i = 1 \text{d}X_{n,i} \frac{1}{h_{2,n}^4} [E_X H_i(X_{n,i}) - EH_i], 1 \leq i \leq n.

Define R_n,i = Y_{n,i} + Z_n,i, and \hat{R} = \frac{1}{n} \sum_{i=1}^n R_n,i where Y_{n,i} is defined as in (46). To prove the central limit theorem (47), it suffices to derive a central limit theorem for \hat{R} - E\hat{f}(t; h_{2,n}) where \hat{R} is the sample mean of i.i.d. random variables R_n,i, 1 \leq i \leq n.

We have

\[ \mathbb{E}\hat{R} = \mathbb{E}R_{n,1} = \mathbb{E}\hat{f}(t; h_{2,n}) = f(t) + h_{2,n}^4 \sum_{|v|=4} \tau_v D_v(1/f)/v! + o(h_{2,n}^4) \]
by Corollary 1 of Giné and Sang [4] and since $\mathbb{E}Z_{n,1} = 0$. Also, $h_{2,n}^d \mathbb{E}Y_{n,1}^2 = \alpha^d(f(t))f(t)\mu_0 + O(h_{2,n}^2)$ by Proposition 4.1 in Section 4. Since

$$h_{2,n}^d \text{Var}(R_{n,1}) = h_{2,n}^d [\mathbb{E}R_{n,1}^2 - (\mathbb{E}R_{n,1})^2]$$

$$= h_{2,n}^d \mathbb{E}Y_{n,1}^2 + h_{2,n}^d \mathbb{E}Z_{n,1}^2 + 2h_{2,n}^d \mathbb{E}(Y_{n,1}Z_{n,1}) - h_{2,n}^d (\mathbb{E}R_{n,1})^2,$$

we need to calculate the limit of terms $h_{2,n}^d \mathbb{E}(Y_{n,1}Z_{n,1})$ and $h_{2,n}^d \mathbb{E}Z_{n,1}^2$. Let $x = uh_{1,n} + x_1$ and $x_1 = t - vh_{2,n}$ be the change of variables. Then,

$$\frac{1}{h_{1,n}^d h_{2,n}^d} \mathbb{E}(\mathbb{E}_X H_t(X, X_1))^2$$

$$= \frac{1}{h_{1,n}^d h_{2,n}^d} \int_{\mathbb{R}^d} \left[ \int_{\mathbb{R}^d} L \left( \frac{t - x}{h_{2,n}} \alpha(f(x)) \right) (\alpha^d)'(f(x))K\left( \frac{x - x_1}{h_{1,n}} \right) f(x)dx \right] f(x_1)dx_1$$

$$= \frac{1}{h_{2,n}^d} \int_{\mathbb{R}^d} \left[ \int_{\mathbb{R}^d} L \left( \frac{t - x - h_{1,n}u}{h_{2,n}} \alpha(f(x + h_{1,n}u)) \right) (\alpha^d)'(f(x + h_{1,n}u))K(u) \right.\times f(x_1 + h_{1,n}u)du \left. \right]^2 f(x_1)dx_1$$

$$= \int_{\mathbb{R}^d} \left[ \int_{\mathbb{R}^d} L \left( \frac{t - h_{2,n}v + h_{1,n}u}{h_{2,n}} \right) \alpha(f(t)) \right. (\alpha^d)'(f(t - h_{2,n}v + h_{1,n}u)) \times K(u) f(t - h_{2,n}v + h_{1,n}u)du \left. \right]^2 f(t - h_{2,n}v)dv$$

$$\xrightarrow{n \rightarrow \infty} \frac{f^3(t)}{\alpha^d(f(t))} \int_{\mathbb{R}^d} L^2(z)dz,$$

and

$$\frac{1}{h_{1,n}^d} \mathbb{E}[Y_{n,1} \mathbb{E}_X (H_t(X, X_1))]$$

$$= \frac{1}{h_{1,n}^d h_{2,n}^d} \int_{\mathbb{R}^d} \left[ \int_{\mathbb{R}^d} L \left( \frac{t - x}{h_{2,n}} \alpha(f(x)) \right) (\alpha^d)'(f(x))K\left( \frac{x - x_1}{h_{1,n}} \right) K\left( \frac{t - x_1}{h_{2,n}} \alpha(f(x_1)) \right) \times (\alpha^d)(f(x_1)) f(x) f(x_1) dx dx_1 \right.$$

$$\times \left. \int_{\mathbb{R}^d} L \left( \frac{u - uh_{1,n}}{h_{2,n}} \right) \alpha(f(uh_{1,n} + t - vh_{2,n})) \right(\alpha^d)'(f(uh_{1,n} + t - vh_{2,n}))K(u) \times K(v\alpha(f(t - h_{2,n}v)))(\alpha^d)(f(t - h_{2,n}v)) f(t - h_{2,n}v) f(uh_{1,n} + t - vh_{2,n}) du dv \right.$$

$$\xrightarrow{n \rightarrow \infty} \frac{1}{2} df^2(t)(\alpha^d)'(f(t))\mu_0$$

since $\int_{\mathbb{R}^d} K(t) \sum_{i=1}^d t_i K'_i(t) dt = -d\mu_0/2$. By the change of variables, it is easy to see that $\mathbb{E}(H_t)$ is bounded by $Ch_{1,n}^d h_{2,n}^d$ for some $C > 0$ and then $\frac{1}{h_{1,n}^d h_{2,n}^d} [\mathbb{E}(H_t)]^2 \rightarrow 0$ as
\[ n \to \infty. \text{ Hence,} \]
\[
 h_{2,n}^d \mathbb{E} Z_{n,1}^2 = \frac{1}{d^2 h_{1,n}^d h_{2,n}^d} \mathbb{E} (\mathbb{E}_X H_t(X, X_1))^2 - \frac{1}{d^2 h_{1,n}^d h_{2,n}^d} \mathbb{E} (H_t)^2
\]
\[
 \xrightarrow{n \to \infty} f^3(t) \left[ \frac{(\alpha^d)'(f(t))}{d^2 \alpha^d(f(t))} \right] \int_{\mathbb{R}^d} L^2(z) dz
\]
and
\[
 h_{2,n}^d \mathbb{E} (Y_{n,1}Z_{n,1}) = \frac{1}{dh_{1,n}^d} \mathbb{E} [Y_{n,1} \mathbb{E}_X (H_t(X, X_1))] - \frac{1}{dh_{1,n}^d} \mathbb{E} Y_{n,1} \mathbb{E}_X H_t
\]
\[
 \xrightarrow{n \to \infty} \frac{1}{2} f^2(t)(\alpha^d)'(f(t)) \mu_0.
\]
Thus,
\[
 h_{2,n}^d \text{Var}(R_{n,1}) \xrightarrow{n \to \infty} \sigma_r^2.
\]
Hence, by central limit theorem for i.i.d. random variables, we have
\[
 \sqrt{nh_{2,n}^d} [\bar{R} - \mathbb{E}\bar{R}] \xrightarrow{D} N(0, \sigma_r^2)
\]
and by Slutsky’s theorem,
\[
 \sqrt{nh_{2,n}^d} [\hat{f}(t; h_{1,n}, h_{2,n}) - \mathbb{E}\hat{f}(t; h_{1,n}, h_{2,n})] \xrightarrow{D} N(0, \sigma_r^2).
\]
Since the term (52) = \[\mathbb{E}\hat{f}(t; h_{1,n}, h_{2,n}) - f(t) = \sum_{|v|=4} \tau_v D_v(1/f) v! h_{2,n}^d (1 + o(1))\] by Proposition 3.1,
\[
 \sqrt{nh_{2,n}^d} [\mathbb{E}\hat{f}(t; h_{1,n}, h_{2,n}) - f(t)] = c_2^{(d+8)/2} \sum_{|v|=4} \tau_v D_v(1/f) v!(1 + o(1)).
\]
Thus,
\[
 \sqrt{nh_{2,n}^d} [\hat{f}(t; h_{1,n}, h_{2,n}) - f(t)] \xrightarrow{D} N \left( c_2^{(d+8)/2} \sum_{|v|=4} \tau_v D_v(1/f) v!, \sigma_r^2 \right).
\]

With the central limit theorem in Theorem 5.1, one can have better statistical inference on the density function value \( f(t) \) at a fixed point \( t \). For example, with some fixed confidence level, the confidence interval for \( f(t) \) using the variable bandwidth kernel estimation is better (in term of the length of the confidence interval) than the classical one since the bandwidth \( h_{2,n} \) here has order of \( n^{-1/(8+d)} \) instead of \( n^{-1/(4+d)} \). The simulation study next in Section 6 demonstrates Theorem 5.1 and shows this advantage of the plug-in variable bandwidth kernel density estimator over the classical one for different distributions.
6 Simulation

In this section we evaluate the performance of the variable bandwidth kernel density estimator (VKDE), (7), in one dimensional case. Instead of the plug-in estimator (7), Jones, McKay and Hu [7] did simulation study for the ideal estimator (2) in one dimensional case. First of all, we provide a result on the integrated mean squared error (IMSE) of the VKDE and therefore a formula of optimal bandwidth.

**Theorem 6.1** Under the conditions in Proposition 3.1 and Theorem 5.1, the IMSE on \( D_r \) is

\[
R(h_{1,n}, h_{2,n})|_{D_r} = h_{2,n}^8 \int_{D_r} \left( \sum_{|v|=4} \tau_v D_v(1/f)/v! \right)^2 dt + \frac{1}{nh_{2,n}^d} \int_{D_r} \sigma_t^2 dt + o(h_{2,n}^8),
\]

Furthermore, the optimal bandwidth \( h_{2,n}^* \) is given by

\[
h_{2,n}^* = \left[ \frac{n \int_{D_r} \left( \sum_{|v|=4} \tau_v D_v(1/f)/v! \right)^2 dt}{\int_{D_r} \sigma_t^2 dt} \right]^{-1/(8+d)}. \tag{57}
\]

**Proof.** From the analysis of Theorem 5.1, it is clear that \( Var(\hat{f}(t; h_{1,n}, h_{2,n})) = (1+o(1)) \sigma_t^2/nh_{2,n}^d \) for \( t \in D_r \). Together with Proposition 3.1, we have (56). The optimal bandwidth (57), which minimize the IMSE, is obvious from the IMSE formula (56).

We compare the performance of VKDE and KDE by conducting one dimensional simulation study of the t-distribution with degree of freedom 4 (\( t_4(0,1) \)), Cauchy(0,1) and Pareto(2,1). The sample size is \( n = 50,000 \) for each simulation study. For all the simulations, we use KDE as in (1) with the normal kernel function. We use the code `density()` in the programming software R and the default bandwidth chosen by R in the estimation for \( t_4(0,1) \). For Cauchy(0,1) or Pareto(2,1), the code `density()` in R can not provide a classical kernel density estimate. Instead, we make new code and select the bandwidth which optimizes the performance among a variety of bandwidths. For VKDE, we assume that \( h_{1,n} = n^{-1/5} \), \( h_{2,n} = n^{-1/9} \), and use the Tricube kernel:

\[
K(u) = \frac{70}{81} (1 - |u|^3)^3 1_{|u| \leq 1}
\]
in either the pilot kernel density estimator or the plug-in estimator (7). The following five time differentiable clipping function \( p \) with \( t_0 = 2 \) (Giné and Sang [4]) is applied:

\[
p(t) = \begin{cases} 
1 + \frac{t^6}{64} \left( 1 - 2(t-2) + \frac{9}{4}(t-2)^2 - \frac{7}{4}(t-2)^3 + \frac{7}{8}(t-2)^4 \right) & \text{if } 0 \leq t \leq 2 \\
t & \text{if } t \geq 2 \\
1 & \text{if } t \leq 0
\end{cases}
\]

The simulation study in Figure 1 shows that, for each of these three distributions, VKDE has better performance than KDE, especially in the tail area.

17
Figure 1: The probability density functions of t-distribution $t_4(0,1)$, Cauchy$(0,1)$ and Pareto$(2,1)$, the kernel density estimates (KDE), and the variable kernel density estimates (VKDE) with 50,000 observations generated from t-distribution $t_4(0,1)$, Cauchy$(0,1)$ and Pareto$(2,1)$ distribution. The left one shows the estimate in the main area with the mode. The right one shows the estimate in the tail area.
We then numerically study the asymptotic distribution of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) and provide the 95% confidence interval for \( f(t) \) at each \( t \). Figures 2 and 3 show the histograms with kernel density fits and normal Q-Q plots of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) at \( t = 1, 3, 5 \) from \( m = 500 \) estimates with \( n = 5,000 \) simulated observations from the t-distribution \( t_4(0, 1) \) for each estimate. The results for the Cauchy(0,1) and Pareto(2,1) distributions at \( t = 1, 3, 5 \) are illustrated in Figures 4, 5 and Figure 6, 7 respectively. It is clear that the estimates of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) are normally distributed in each case, especially when \( t \) is not very big and \( f(t) \) is not very small. It is evident that the distribution of the estimates \( \hat{f}(t; h_{1,n}, h_{2,n}) \) at \( t = 5 \) in the \( t_4(0, 1) \) case is slightly away from the normal distribution since the \( t_4(0, 1) \) distribution has relatively lighter tail than the other two distributions and the central limit theorem in Theorem 5.1 only holds for \( t \in \mathcal{D}_r \). We also performed simulation study at \( t = 0, 2, 4, 6 \) in the \( t_4(0, 1) \) and Cauchy(0,1) cases and at \( t = 2, 4, 6 \) in the Pareto(2,1) case and obtained similar plots and results. Over all, the simulation study here confirms Theorem 5.1.

![Histogram: estimates of \( f(1) \)](image)

![Histogram: estimates of \( f(3) \)](image)

![Histogram: estimates of \( f(5) \)](image)

Figure 2: The histograms with kernel density fits of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) at \( t = 1, 3, 5 \) from \( m = 500 \) estimates with \( n = 5,000 \) simulated observations from the t-distribution \( t_4(0, 1) \) for each estimate.
Figure 3: The normal Q-Q plots of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) at \( t = 1, 3, 5 \) from \( m = 500 \) estimates with \( n = 5,000 \) simulated observations from the \( t \)-distribution \( t_4(0, 1) \) for each estimate.

Figure 4: The histograms with kernel density fits of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) at \( t = 1, 3, 5 \) from \( m = 500 \) estimates with \( n = 5,000 \) simulated observations from Cauchy(0,1) distribution for each estimate.
Figure 5: The normal Q-Q plots of $\hat{f}(t; h_{1,n}, h_{2,n})$ at $t = 1, 3, 5$ from $m = 500$ estimates with $n = 5,000$ simulated observations from Cauchy(0,1) distribution for each estimate.

Figure 6: The histograms with kernel density fits of $\hat{f}(t; h_{1,n}, h_{2,n})$ at $t = 1, 3, 5$ from $m = 500$ estimates with $n = 5,000$ simulated observations from Pareto(2,1) distribution for each estimate.
Figure 7: The normal Q-Q plots of \( \hat{f}(t; h_{1,n}, h_{2,n}) \) at \( t = 1, 3, 5 \) from \( m = 500 \) estimates with \( n = 5,000 \) simulated observations from Pareto(2,1) distribution for each estimate.

Tables 1, 2 and 3 provide confidence intervals and their widths for the densities at different points. KDE ll and KDE ul are the lower and upper limits of the confidence interval generated by the classical kernel density estimator. VKDE ll and VKDE ul are the lower and upper limits of the confidence interval generated by the plug-in variable bandwidth kernel density estimator. In each case 500 estimates are used to generate the confidence interval and for each estimate the sample size is \( n = 5,000 \). The results show that each confidence interval covers the corresponding true density value \( f(t) \) and the confidence interval generated by the plug-in variable bandwidth kernel density estimator is always shorter than the confidence interval generated by the classical kernel density estimator in each case. This shows the advantage of the variable bandwidth kernel method over the classical kernel method.

Table 1: The 95% confidence intervals for density \( f(t) \) of the t-distribution \( t_4(0,1) \) at values \( t = 0, 1, 2, 3, 4, 5, 6 \).

<table>
<thead>
<tr>
<th>( t )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(t) )</td>
<td>3.750e-1</td>
<td>2.147e-1</td>
<td>6.629e-2</td>
<td>1.969e-2</td>
<td>6.708e-3</td>
<td>2.650e-3</td>
<td>1.186e-3</td>
</tr>
<tr>
<td>KDE ll</td>
<td>3.721e-1</td>
<td>2.144e-1</td>
<td>6.490e-2</td>
<td>1.911e-2</td>
<td>6.615e-3</td>
<td>2.364e-3</td>
<td>1.109e-3</td>
</tr>
<tr>
<td>KDE ul</td>
<td>3.764e-1</td>
<td>2.176e-1</td>
<td>6.667e-2</td>
<td>2.009e-2</td>
<td>7.166e-3</td>
<td>2.720e-3</td>
<td>1.356e-3</td>
</tr>
<tr>
<td>width</td>
<td>4.269e-3</td>
<td>3.182e-3</td>
<td>1.772e-3</td>
<td>9.810e-4</td>
<td>5.511e-4</td>
<td>3.563e-4</td>
<td>2.471e-4</td>
</tr>
<tr>
<td>VKDE ll</td>
<td>3.727e-1</td>
<td>2.145e-1</td>
<td>6.559e-2</td>
<td>1.929e-2</td>
<td>6.569e-3</td>
<td>2.568e-3</td>
<td>1.207e-3</td>
</tr>
<tr>
<td>VKDE ul</td>
<td>3.767e-1</td>
<td>2.173e-1</td>
<td>6.674e-2</td>
<td>1.976e-2</td>
<td>6.781e-3</td>
<td>2.676e-3</td>
<td>1.268e-3</td>
</tr>
<tr>
<td>width</td>
<td>4.048e-3</td>
<td>2.788e-3</td>
<td>1.151e-3</td>
<td>4.729e-4</td>
<td>2.113e-4</td>
<td>1.078e-4</td>
<td>6.113e-5</td>
</tr>
</tbody>
</table>
Table 2: The 95% confidence intervals for density $f(t)$ of the Cauchy distribution $\text{Cauchy}(0,1)$ at values $t = 0, 1, 2, 3, 4, 5, 6$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(t)$</td>
<td>3.183e-1</td>
<td>1.592e-1</td>
<td>6.366e-2</td>
<td>3.183e-2</td>
<td>1.872e-2</td>
<td>1.224e-2</td>
<td>8.603e-3</td>
</tr>
<tr>
<td>KDE ll</td>
<td>3.159e-1</td>
<td>1.584e-1</td>
<td>6.291e-2</td>
<td>3.133e-2</td>
<td>1.855e-2</td>
<td>1.182e-2</td>
<td>8.339e-3</td>
</tr>
<tr>
<td>KDE ul</td>
<td>3.197e-1</td>
<td>1.611e-1</td>
<td>6.465e-2</td>
<td>3.276e-2</td>
<td>1.952e-2</td>
<td>1.262e-2</td>
<td>8.975e-3</td>
</tr>
<tr>
<td>width</td>
<td>3.733e-3</td>
<td>2.687e-3</td>
<td>1.736e-3</td>
<td>1.227e-3</td>
<td>9.718e-4</td>
<td>7.957e-4</td>
<td>6.360e-4</td>
</tr>
<tr>
<td>VKDE ll</td>
<td>3.173e-1</td>
<td>1.585e-1</td>
<td>6.313e-2</td>
<td>3.148e-2</td>
<td>1.863e-2</td>
<td>1.204e-2</td>
<td>8.467e-3</td>
</tr>
<tr>
<td>width</td>
<td>3.413e-3</td>
<td>2.072e-3</td>
<td>1.146e-3</td>
<td>6.830e-4</td>
<td>4.755e-4</td>
<td>3.369e-4</td>
<td>2.524e-4</td>
</tr>
</tbody>
</table>

Table 3: The 95% confidence intervals for density $f(t)$ of the Pareto distribution $\text{Pareto}(2,1)$ at values $t = 1, 2, 3, 4, 5, 6$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(t)$</td>
<td>2.5e-1</td>
<td>7.407e-2</td>
<td>3.125e-2</td>
<td>1.6e-2</td>
<td>9.259e-3</td>
<td>5.831e-3</td>
</tr>
<tr>
<td>KDE ll</td>
<td>2.493e-1</td>
<td>7.348e-2</td>
<td>3.024e-2</td>
<td>1.559e-2</td>
<td>9.111e-3</td>
<td>5.578e-3</td>
</tr>
<tr>
<td>width</td>
<td>3.316e-3</td>
<td>1.913e-3</td>
<td>1.212e-3</td>
<td>8.780e-4</td>
<td>6.250e-4</td>
<td>5.280e-4</td>
</tr>
<tr>
<td>VKDE ll</td>
<td>2.493e-1</td>
<td>7.359e-2</td>
<td>3.091e-2</td>
<td>1.591e-2</td>
<td>9.201e-3</td>
<td>5.765e-3</td>
</tr>
<tr>
<td>width</td>
<td>2.056e-3</td>
<td>8.880e-4</td>
<td>4.410e-4</td>
<td>2.740e-4</td>
<td>1.920e-4</td>
<td>1.370e-4</td>
</tr>
</tbody>
</table>

Acknowledgement The authors thank the referees, the Associate Editor and the Editor for their careful reading of the manuscript and for their insightful comments, which have helped to improve the quality of this paper.

References


