Uniform Consistency of a Boundary Corrected Kernel Density Estimator

Huihui Li

Department of Economics, Pennsylvania State University

October 2015

Abstract

Zhang, Karunamuni, and Jones (1999) proposed a method of boundary correction for kernel density estimation, which is later improved by Karunamuni and Zhang (2008). This method uses the reflection technique involving reflecting a transformation of the data. In this paper, I consider a generalization of Zhang, Karunamuni, and Jones's estimator. The generalized estimator allows one to consistently estimate the density function, not only on a compact support, but also on arbitrary compact subinterval of the support on which the density is continuous but possibly has discontinuity at the endpoints. I establish the uniform consistency of the generalized estimator and show that it has a uniform convergence rate of $O_p\left(h^2 + \sqrt{\log n/(nh)}\right)$, provided that the primary and the secondary bandwidths shrink at proper rates. The potential extension of the estimator in order to correct higher order bias is also discussed in the paper.

Keywords: Kernel density estimation, boundary effects, uniform convergence, pseudodata, transformation.

JEL Classification: C13, C14

1 Introduction

Let f denote a probability density function with support $\mathbb{X}$, and consider a random sample $X_1, \ldots, X_n \in \mathbb{R}$ drawn from f. The standard kernel estimator of f at $x \in \mathbb{X}$ is given by

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where h is the bandwidth and K is a kernel function that satisfies $\int_{-\infty}^{\infty} K(u)\,\mathrm{d}u = 1$, $\int_{-\infty}^{\infty} uK(u)\,\mathrm{d}u = 0$ and $\int_{-\infty}^{\infty} u^2K(u)\,\mathrm{d}u < \infty$. When the support $\mathbb{X}$ is bounded then the standard kernel density estimator (1.1) is not necessarily consistent at those points near the boundary of the support unless $f(x) = 0$. This phenomenon is known as the "boundary effect" of the standard kernel estimator. If the density f is discontinuous at some points in the interior of the support, similar problem can also happen around these points of discontinuity and the consistency of standard kernel estimator will break down, also.

It has already been extensively studied in literature how to correct this boundary effect. Some well-known methods include the simple reflection method (Cline and Hart, 1991, Schuster, 1985), the boundary kernel method (Gasser and Müller, 1979, Gasser, Müller, and Mammitzsch, 1985, Müller, 1991), the local linear method (Cheng, 1994, Cheng, Fan, and Marron, 1997), the transformation method (Marron and Ruppert, 1994, Wand, Marron, and Ruppert, 1991), and the pseudodata method (Cowling and Hall, 1996). Zhang, Karunamuni, and Jones (1999, hereforth ZKJ for short) proposed an improved boundary corrected density estimator, which is a combination of the pseudodata, transformation, and reflection methods. Their method consists of three basic steps: first, use some function g to transform the original data $X_1, \ldots, X_n$ to a new set of observations while keeping the original data; second, reflect those transformed data around the boundary of the support; finally, estimate the density function in the same way as standard kernel estimator (1.1) but based on the enlarged data sample including the original data as well as the transformed and reflected data. I will focus on the ZKJ method because this method has several advantages compared with other boundary correction approaches. Whereas the simple reflection has bad bias and those approaches involving only kernel modifications such as the boundary kernel-related methods are usually associated with larger variance, the ZKJ estimator both controls the bias and keeps the variance relatively small. This makes the ZKJ estimator present better performance for various shapes of densities. Unlike the boundary kernel method and the local linear method, the ZKJ method ensures that the estimated density function is nonnegative everywhere. Moreover, the ZKJ method only corrects the boundary bias in the region where the standard kernel estimator loses consistency and as a result the ZKJ estimator reduces to the standard kernel estimator numerically in the interior region which is at least one bandwidth distance away from the boundary points. These features allow the ZKJ estimator to inherit all the good properties of the standard kernel estimator in the interior region and make the ZKJ estimator easy to compute.

Zhang, Karunamuni, and Jones (1999) and Karunamuni and Zhang (2008) give the pointwise asymptotic properties of the ZKJ estimator. However, in many applications such as estimating economic models, researchers are more interested in knowing the whole density function and this requires that the estimator converges to the true density uniformly over the inference region. To address this problem, this paper will establishes the uniform consistency of the ZKJ estimator by showing the uniform rate of convergence. In Section 2, I formally define the estimator under consideration and show its uniform rate of convergence. I will briefly discuss an extension of the ZKJ method—higher-order bias reduction—and conclude my study in Section 3. Proofs are deferred to the Appendix.

2 Boundary-corrected kernel density estimator and uniform consistency

For a random sample $X_1, \ldots, X_n \in \mathbb{R}$ from an unknown density f with support $\mathbb{X}$, following Zhang, Karunamuni, and Jones (1999), I define the *boundary-corrected kernel density estimator on compact interval* $[a, b] \subseteq \mathbb{X}$ as

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x - a + g_1(X_i - a)}{h}\right) + K\left(\frac{b - x + g_2(b - X_i)}{h}\right) \right], \quad (2.1)$$

where

$$g_1(u) = u + d_1 u^2 + A_1 d_1^2 u^3, \quad g_2(u) = u + d_2 u^2 + A_2 d_2^2 u^3 \quad (2.2)$$

are transformation functions with

$$d_1 = \frac{f'(a)}{f(a)}, \quad d_2 = -\frac{f'(b)}{f(b)}.$$

Here the kernel function K, bandwidth h, and the constants $A_1, A_2 > \frac{1}{3}$ are all tuning parameters.

The density estimator (2.1) is slightly different from the one in Zhang, Karunamuni, and Jones (1999). My estimator essentially estimate the density on an interval by using only those observations lying in that interval and the boundary correction is implemented in form at both endpoints. I define the estimator in the form for the purpose of versatility. My estimator can handle, not only the standard problem of correcting boundary effect at the boundaries of support $\mathbb{X}$ by simply setting $[a, b] = \mathbb{X}$, but also the inconsistency issue raised by the discontinuity of density function. The latter is achieved by following similar approach suggested by Schuster (1985) and Cline and Hart (1991), in which I let one or both of a, b be the discontinuous point(s) of f so that f is continuous inside interval $[a, b]$.[1]

Note that g_1 and g_2 depend on the unknown density function then the kernel density estimator $\tilde{f}$ given in (2.1) is usually infeasible unless d_1 and d_2 are known. Thus, I define the feasible counterpart of $\tilde{f}$ as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) \left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x - a + \hat{g}_1(X_i - a)}{h}\right) + K\left(\frac{b - x + \hat{g}_2(b - X_i)}{h}\right) \right], \quad (2.3)$$

where $\hat{g}_1$ and $\hat{g}_2$ are obtained by replacing d_1 and d_2 in (2.2) with the following estimators

$$\hat{d}_1 = \frac{1}{h_1} \left\{ \log \left[\frac{1}{nh_1} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) K\left(\frac{h_1 - X_i + a}{h_1}\right) + \frac{1}{n^2} \right] - \log \left[\max \left(\frac{1}{nh_0} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) K_0\left(\frac{a - X_i}{h_0}\right), \frac{1}{n^2} \right) \right] \right\}, \quad (2.4)$$

$$\hat{d}_2 = \frac{1}{h_1} \left\{ \log \left[\frac{1}{nh_1} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) K\left(\frac{h_1 + X_i - b}{h_1}\right) + \frac{1}{n^2} \right] - \log \left[\max \left(\frac{1}{nh_0} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) K_0\left(\frac{X_i - b}{h_0}\right), \frac{1}{n^2} \right) \right] \right\}, \quad (2.5)$$

where $K_0(u)$ is a so-called endpoint kernel satisfying

$$\int_{-1}^{0} K_0(u)\, \mathrm{d}u = 1, \quad \int_{-1}^{0} u K_0(u)\, \mathrm{d}u = 0, \quad \int_{-1}^{0} u^2 K_0(u)\, \mathrm{d}u \neq 0,$$

and

$$h_0 = h_1 \cdot \left[\frac{\left(\int_{-1}^{1} u^2 K(u)\, \mathrm{d}u\right)^2 \left(\int_{-1}^{0} K_0^2(u)\, \mathrm{d}u\right)}{\left(\int_{-1}^{0} u^2 K_0(u)\, \mathrm{d}u\right)^2 \left(\int_{-1}^{1} K^2(u)\, \mathrm{d}u\right)} \right]^{1/5}. \quad (2.6)$$

Specifically, I can set $K_0(u) = (6 + 18u + 12u^2) \cdot I(-1 \leqslant u \leqslant 0)$ which is the optimal endpoint kernel in the sense of minimizing the MSE at the boundary point among all the kernel functions that change sign only once on the support $[-\infty, 0]$ and satisfy $\int_{-\infty}^{0} K(u)\, \mathrm{d}u = 1$ as well as $\int_{-\infty}^{0} uK(u)\, \mathrm{d}u = 0$ (see Zhang and Karunamuni, 1998). Following Karunamuni and Zhang (2008), I allow the bandwidth h_1 (secondary band-

[1]In fact, the results of this paper keep valid even if f is smooth at a and b but the boundary correction is forcibly implemented. However, this is not recommended since it unnecessarily causes poor performance of the density estimates.

width) used in estimating d_1, d_2 to differ from the bandwidth h (primary bandwidth) used in estimating f. This will bring potential improvement of the estimator's performance when the uniform consistency is of interest.

2.1 Uniform consistency of the infeasible estimator

I start with a series of assumptions about the underlying data generating process and the choice of tuning parameter values.

Assumption A. *X_i, $i = 1, 2, \ldots, n$ are independently and identically distributed as f.*

Assumption B. *f is twice continuously differentiable with bounded f'' on $[a, b]$. $f(x) \geqslant c_0 > 0$ for all $x \in [a, b]$.*

Assumption C. *The kernel K is symmetric with support $[-1, 1]$ and has twice continuous bounded derivative on $\mathbb{R}$. K is of order 2, i.e. $\int_{-\infty}^{\infty} K(u)\,\mathrm{d}u = 1$, $\int_{-\infty}^{\infty} uK(u)\,\mathrm{d}u = 0$, $\int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u = \kappa < \infty$.*

Assumption D. *h satisfies $0 < h < (b-a)/2$. $h \to 0$ and $nh/\log n \to \infty$ as $n \to \infty$.*

Now I show the uniform consistency of the infeasible estimator (2.1) on $[a, b]$.

Lemma 1. *Under Assumptions A, B, C and D, when n is sufficiently large,*

$$\sup_{x \in [a,b]} \left| E\tilde{f}(x) - f(x) \right| = O(h^2). \tag{2.7}$$

Proof. See Appendix A.1. □

Lemma 2. *Under Assumptions A, B, C and D, when n is sufficiently large,*

$$\sup_{x \in [a,b]} \mathrm{Var}\left(\tilde{f}(x)\right) = O\left(\frac{1}{nh}\right). \tag{2.8}$$

Proof. See Appendix A.2. □

Theorem 1. *Under Assumptions A, B, C and D, as $n \to \infty$,*

$$\sup_{x \in [a,b]} |\tilde{f}(x) - f(x)| = O\left(h^2 + \sqrt{\frac{\log n}{nh}}\right) \quad \textit{almost surely.} \tag{2.9}$$

Proof. Lemmas 1 and 2 give the uniform convergence rates of the bias and the variance of estimator (2.1). Given those, the proof is similar to showing the uniform rate of convergence for standard kernel estimator on a compact interior subset[2] of the support (see, e.g. Li and Racine, 2006, section 1.12) and is therefore omitted here. □

[2] A set A is called an interior subset of set B if it is a subset of the interior of B.

2.2 Uniform consistency of the feasible estimator

Now I turn to discuss the uniform convergence rate of the feasible estimator (2.3). I denote the corresponding terms in (2.4) and (2.5) by

$$f_1^*(a+h_1) \equiv \frac{1}{nh_1}\sum_{i=1}^n I(a \leqslant X_i \leqslant b) K\left(\frac{h_1 - X_i + a}{h_1}\right), \quad f_1^*(a) \equiv \frac{1}{nh_0}\sum_{i=1}^n I(a \leqslant X_i \leqslant b) K_0\left(\frac{a - X_i}{h_0}\right),$$
$$f_2^*(b-h_1) \equiv \frac{1}{nh_1}\sum_{i=1}^n I(a \leqslant X_i \leqslant b) K\left(\frac{h_1 + X_i - b}{h_1}\right), \quad f_2^*(b) \equiv \frac{1}{nh_0}\sum_{i=1}^n I(a \leqslant X_i \leqslant b) K_0\left(\frac{X_i - b}{h_0}\right),$$

and

$$f_1(a+h_1) \equiv f_1^*(a+h_1) + \frac{1}{n^2}, \qquad f_1(a) \equiv \max\left(f_1^*(a), \frac{1}{n^2}\right),$$
$$f_2(b-h_1) \equiv f_1^*(b-h_1) + \frac{1}{n^2}, \qquad f_2(b) \equiv \max\left(f_2^*(b), \frac{1}{n^2}\right).$$

Here $f_1(\cdot)$, $f_1^*(\cdot)$, $f_2(\cdot)$ and $f_2^*(\cdot)$ can be seen as some kind of estimators of f at the corresponding points, and then $\hat{d}_1$ and $\hat{d}_2$ can be written as

$$\hat{d}_1 = \frac{\log f_1(a+h_1) - \log f_1(a)}{h_1}, \qquad \hat{d}_2 = \frac{\log f_2(b-h_1) - \log f_2(b)}{h_1}. \tag{2.10}$$

I first show some consistency results for f_1^*, f_2^*.

Lemma 3. *Under Assumptions A, B and C, if $h_1 \to 0$ and $nh_1 \to \infty$ as $n \to \infty$, then for sufficiently large n and any $1 \leqslant p \leqslant n$,*

$$E\left\{[f_1^*(a+h_1) - f(a+h_1)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{nh_1}\right), \tag{2.11}$$
$$E\left\{[f_1^*(a) - f(a)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{n^2h_1}\right), \tag{2.12}$$
$$E\left\{[f_2^*(b-h_1) - f(b-h_1)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{nh_1}\right), \tag{2.13}$$
$$E\left\{[f_2^*(b) - f(b)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{nh_1}\right). \tag{2.14}$$

Proof. See Appendix A.3. □

Because the factor $1/n^2$ in (2.4) and (2.5) is used to make f_1 and f_2 bounded away from zero, in fact, it does not affect the statistical properties of f_1^* and f_2^*. Thus, the following lemma directly follows Lemma 3.

Lemma 4. *Under Assumptions A, B and C, if $h_1 \to 0$ and $nh_1 \to \infty$ as $n \to \infty$, then for sufficiently large n and any $1 \leqslant p \leqslant n$,*

$$E\left\{[f_1(a+h_1) - f(a+h_1)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{nh_1}\right), \tag{2.15}$$
$$E\left\{[f_1(a) - f(a)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{nh_1}\right), \tag{2.16}$$

$$E\left\{[f_2(b-h_1)-f(b-h_1)]^2 \mid X_p\right\} = O\left({h_1}^4+\frac{1}{nh_1}\right), \tag{2.17}$$

$$E\left\{[f_2(b)-f(b)]^2 \mid X_p\right\} = O\left({h_1}^4+\frac{1}{nh_1}\right). \tag{2.18}$$

Proof. See Appendix A.4. □

The main intermediate step in proving the uniform convergence rate of the feasible estimator is to establish the statistical properties of the estimator for d_1 and d_2.

Lemma 5. *Under Assumptions A, B and C, if $h_1 \to 0$ and $n{h_1}^3 \to \infty$ as $n \to \infty$, then for sufficiently large n and any $1 \leqslant p \leqslant n$,*

$$E\left[(\hat{d}_1-d_1)^2 \,\middle|\, X_p\right] = O\left({h_1}^2+\frac{1}{n{h_1}^3}\right), \quad E\left[(\hat{d}_2-d_2)^2 \,\middle|\, X_p\right] = O\left({h_1}^2+\frac{1}{n{h_1}^3}\right). \tag{2.19}$$

Proof. See Appendix A.5. □

Lemma 5 implies the "optimal" secondary bandwidth h_1 in the sense of making the convergence rates given by Lemma 5 as fast as possible, is of the form $h_1 = \lambda' n^{-1/5}$. Given that, $E[(\hat{d}_1-d_1)^2 \mid X_p] = O(n^{-2/5})$ and $E[(\hat{d}_2-d_2)^2 \mid X_p] = O(n^{-2/5})$. With the following additional assumption about the convergence rate of h_1 relative to the primary bandwidth h, I can show the feasible density estimator $\hat{f}$ has the same uniform convergence rate (in probability) as the infeasible estimator $\tilde{f}$.

Assumption E. *h_1 satisfies $h_1 = O(h)$ and $1/\sqrt{n{h_1}^3} = O(h)$ as $n \to \infty$.*

And the main result of this paper is the following theorem.

Theorem 2. *Under Assumptions A, B, C, D and E, when n is sufficiently large,*

$$\sup_{x\in[a,b]}\left|\hat{f}(x)-f(x)\right| = O_p\left(h^2+\sqrt{\frac{\log n}{nh}}\right). \tag{2.20}$$

Proof. See Appendix A.6. □

Assumption E requires that as $n \to \infty$, the secondary bandwidth h_1 should converge to zero not slower than the primary bandwidth h but, meanwhile, it should not converge too fast compared to h. When h takes the form of $\lambda n^{-1/5}$, the only possible convergence rate for h_1 that satisfies Assumption E is $h_1 = \lambda' n^{-1/5}$, that is, both h and h_1 have the same convergence rate to zero. And if I take $h = o(n^{-1/5})$, then any sequence of h_1 will violate the condition of Assumption E. Therefore, to make the result in Theorem 2 hold, Assumption E implicitly imposes some restrictions on the primary bandwidth that can be chosen.

It is worth to note that when the support $\mathbb{X}$ is a compact interval and $[a,b] = \mathbb{X}$, Theorem 2 shows that after boundary correction, my estimator can achieve the same uniform rate of convergence on the entire support as what the standard kernel estimator gets on fixed compact interior subset of the support. Provided the assumption that $h = \lambda(\log n/n)^{-1/5}$, my estimator has the optimal uniform convergence rate of order $O_p\left((\log n/n)^{-2/5}\right)$.

3 Conclusion

In this paper, I establish the uniform consistency of a boundary corrected kernel density estimator proposed by Zhang, Karunamuni, and Jones (1999). My results show that under almost the same assumptions, the boundary corrected kernel estimator can extend the existing uniform consistency result of the standard kernel estimator on a fixed compact interior subset of the support to the entire support. More specifically, I show that my boundary corrected estimator uniformly converges to the true density at the rate of $O_p(h^2 + \sqrt{\log n/n})$, and the optimal uniform rate of convergence is $O_p\left((\log n/n)^{-2/5}\right)$ achieved when the bandwidth is chosen in the form of $h = \lambda(\log n/n)^{-1/5}$.

In respect to removing the boundary bias that causes inconsistency, I precisely stick to the method given by Zhang, Karunamuni, and Jones (1999). Their method, like many other boundary correction methods in literature, aims to reduce the bias of the proposed estimator to the order of h^2, which yields a uniform rate of convergence in the order of $O_p(h^2 + \sqrt{\log n/n})$. It is appealing to ask whether the bias can be further reduced, for example, to the order of h^3 or even h^4, by using a similar boundary correction technique, if the underlying density function is smooth enough and it is allowed to choose kernel function freely. For ZKJ method, the key step of boundary correction is done by using a transformation function g to generated transformed data. So following the same idea, it is expected that the higher order bias can be removed if the transformation function is designed appropriately.

For simplicity, let me assume the support $\mathbb{X} = [0, \infty)$ so 0 is the only boundary point, and define the (infeasible) density estimator on $\mathbb{X}$ as

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{n}\left[K\left(\frac{x - X_i}{h}\right) + K\left(\frac{x + g(X_i)}{h}\right)\right],$$

where g is the transformation function. Through studying the expansion of the estimator's pointwise bias, I can find some clues. As ZKJ method does, to reduce the bias to the order of h^2, I need to use a kernel function K at least of order 2 and the transformation function g should satisfy (i) $g(0) = 0$, (ii) g is strictly increasing and $\lim_{u\to\infty} g(u) = \infty$, (iii) $g'(0) = 1$, and (iv) $2f'(0) - g''(0)f(0) = 0$. If I want to reduce the bias to the order of h^3, I need K at least of order 3 and g satisfying conditions (i)–(iv) and (v) $g'''(0)f(0) + 3g''(0)f'(0) - 3g''(0)^2 f(0) = 0$. It is easy to verify that the transformation function $g(u) = u + du^2 + d^2u^3 + Bu^4$ with $d = f'(0)/f(0)$ and $B > 0$ will work. If I want to reduce the bias further, for example, to make the bias have order of h^4, then K has to be at least of order 4 and g must satisfy (vi) $2f'''(0) - 6f''(0)g''(0) + 15f'(0)g''(0)^2 - 15f(0)g''(0)^3 - [4f'(0) - 10f(0)g''(0)]g'''(0) - f(0)g^{(4)}(0) = 0$ in addition to conditions (i)–(v), where $g^{(4)}$ is the fourth derivative of g. Hence, one can see that using this technique to get higher-order bias reduction and boundary correction at the same time is possible. However, the higher order I want to reduce the bias to, the more information about the shape of unknown density f at the boundary point is required. How to invent a valid and feasible estimator, which will inevitably involve pilot estimation of $f(0)$, $f'(0)$, $f''(0)$, etc, is still a question open to answer.

References

CHENG, M.-Y. (1994): "On Boundary Effects of Smooth Curve Estimators," Mimeo Series #2319, Institute of Statistics, University of North Carolina.

CHENG, M.-Y., J. FAN, AND J. S. MARRON (1997): "On Automatic Boundary Corrections," *Annals of Statistics*, 25(4), 1691–1708.

CLINE, D. B. H., AND J. D. HART (1991): "Kernel Estimation of Densities with Discontinuities or Discontinuous Derivatives," *Statistics*, 22(1), 69–84.

COWLING, A., AND P. HALL (1996): "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation," *Journal of the Royal Statistical Society, Series B*, 58(3), 551–563.

GASSER, T., AND H.-G. MÜLLER (1979): "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation*, ed. by T. Gasser, and M. Rosenblatt, pp. 23–68. Springer Berlin Heidelberg.

GASSER, T., H.-G. MÜLLER, AND V. MAMMITZSCH (1985): "Kernels for Nonparametric Curve Estimation," *Journal of the Royal Statistical Society, Series B (Methodological)*, 47(2), 238–252.

KARUNAMUNI, R. J., AND S. ZHANG (2008): "Some Improvements on a Boundary Corrected Kernel Density Estimator," *Statistics and Probability Letters*, 78(5), 499–507.

LI, Q., AND J. S. RACINE (2006): *Nonparametric Econometrics: Theory and Practice*, Economics Books. Princeton University Press.

MARRON, J. S., AND D. RUPPERT (1994): "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society, Series B (Methodological)*, 56(4), 653–671.

MÜLLER, H.-G. (1991): "Smooth Optimum Kernel Estimators Near Endpoints," *Biometrika*, 78(3), 521–530.

SCHUSTER, E. F. (1985): "Incorporating Support Constraints Into Nonparametric Estimators of Densities," *Communications in Statistics - Theory and Methods*, 14(5), 1123–1136.

WAND, M. P., J. S. MARRON, AND D. RUPPERT (1991): "Transformations in Density Estimation," *Journal of the American Statistical Association*, 86(414), 343–353.

ZHANG, S., AND R. J. KARUNAMUNI (1998): "On Kernel Density Estimation Near Endpoints," *Journal of Statistical Planning and Inference*, 70(2), 301–316.

ZHANG, S., R. J. KARUNAMUNI, AND M. C. JONES (1999): "An Improved Estimator of the Density Function at the Boundary," *Journal of the American Statistical Association*, 94(448), 1231–1241.

A Proofs

A.1 Proof of Lemma 1

Since X_i, $i = 1, 2, \ldots, n$ are i.i.d. draws from f by Assumption A, then

$$E\tilde{f}(x) = \frac{1}{h}\int_a^b K\left(\frac{x-z}{h}\right) f(z)\,\mathrm{d}z + \frac{1}{h}\int_a^b K\left(\frac{x-a+g_1(z-a)}{h}\right) f(z)\,\mathrm{d}z + \frac{1}{h}\int_a^b K\left(\frac{b-x+g_2(b-z)}{h}\right) f(z)\,\mathrm{d}z. \quad \text{(A.1)}$$

By changing variable and taking Taylor expansion with Lagrangian remainder, I have

$$
\begin{aligned}
\frac{1}{h}\int_a^b K\left(\frac{x-z}{h}\right) f(z)\,\mathrm{d}z &= \int_{(x-b)/h}^{(x-a)/h} K(t) f(x-ht)\,\mathrm{d}t \\
&= f(x)\int_{(x-b)/h}^{(x-a)/h} K(t)\,\mathrm{d}t - hf'(x)\int_{(x-b)/h}^{(x-a)/h} tK(t)\,\mathrm{d}t \\
&\quad + \frac{h^2}{2}\int_{(x-b)/h}^{(x-a)/h} t^2 K(t) f''(x-\lambda_{0t}ht)\,\mathrm{d}t,
\end{aligned}
\tag{A.2}
$$

where $\lambda_{0t} \in [0,1]$ may depend on t. Using the properties that $g_1(0) = 0$, $g_1'(0) = 1$ and $g_1''(0) = 2f'(a)/f(a)$,[3] I have

$$
\begin{aligned}
\frac{1}{h}\int_a^b K\left(\frac{x-a+g_1(z-a)}{h}\right) f(z)\,\mathrm{d}z &= \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\frac{f(a+g_1^{-1}(ht-x+a))}{g_1'(g_1^{-1}(ht-x+a))}\,\mathrm{d}t \\
= \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\Bigg\{ &\frac{f(a+g_1^{-1}(0))}{g_1'(g_1^{-1}(0))} \\
&+ h\left(t-\frac{x-a}{h}\right)\frac{f'(a+g_1^{-1}(0))g_1'(g_1^{-1}(0)) - f(a+g_1^{-1}(0))g_1''(g_1^{-1}(0))}{[g_1'(g_1^{-1}(0))]^3} \\
&+ \frac{h^2}{2}\left(t-\frac{x-a}{h}\right)^2 \mathcal{G}_1(\lambda_{1t}(ht-x+a))\Bigg\}\,\mathrm{d}t \\
= [f(a)+(x-a)f'(a)]\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} &K(t)\,\mathrm{d}t - hf'(a)\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} tK(t)\,\mathrm{d}t \\
&+ \frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h}\left(t-\frac{x-a}{h}\right)^2 K(t)\mathcal{G}_1(\lambda_{1t}(ht-x+a))\,\mathrm{d}t
\end{aligned}
\tag{A.3}
$$

where $\lambda_{1t} \in [0,1]$ may depend on t and

$$
\begin{aligned}
\mathcal{G}_1(u) = &\frac{f''(a+g_1^{-1}(u))g_1'(g_1^{-1}(u)) - f(a+g_1^{-1}(u))g_1'''(g_1^{-1}(u))}{[g_1'(g_1^{-1}(u))]^4} \\
&- \frac{3g_1''(g_1^{-1}(u))[f'(a+g_1^{-1}(u))g_1'(g_1^{-1}(u)) - f(a+g_1^{-1}(u))g_1''(g_1^{-1}(u))]}{[g_1'(g_1^{-1}(u))]^5}.
\end{aligned}
$$

Similarly, using the symmetry of K and the properties that $g_2(0) = 0$, $g_2'(0) = 1$ and $g_2''(0) = -2f'(b)/f(b)$, I have

$$
\begin{aligned}
\frac{1}{h}\int_a^b K\left(\frac{b-x+g_2(b-z)}{h}\right) f(z)\,\mathrm{d}z &= [f(b)+(x-b)f'(b)]\int_{[x-b-g_2(b-a)]/h}^{(x-b)/h} K(t)\,\mathrm{d}t \\
- hf'(b)\int_{[x-b-g_2(b-a)]/h}^{(x-b)/h} tK(t)\,\mathrm{d}t &+ \frac{h^2}{2}\int_{[x-b-g_2(b-a)]/h}^{(x-b)/h}\left(t-\frac{x-b}{h}\right)^2 K(t)\mathcal{G}_2(\lambda_{2t}(-ht+x-b))\,\mathrm{d}t
\end{aligned}
\tag{A.4}
$$

where $\lambda_{2t} \in [0,1]$ may depend on t and

[3] By construction, g_1 and g_2 are strictly increasing continuous functions satisfying $g_1(0) = g_2(0) = 0$, $g_1'(0) = g_2'(0) = 1$, $g_1''(0) = 2f'(a)/f(a)$ and $g_2''(0) = -2f'(b)/f(b)$.

$$\mathcal{G}_2(u) = \frac{f''(b - g_2^{-1}(u))g_2'(g_2^{-1}(u)) - f(b - g_2^{-1}(u))g_2'''(g_2^{-1}(u))}{[g_2'(g_2^{-1}(u))]^4} - \frac{3g_2''(g_2^{-1}(u))[f'(b - g_2^{-1}(u))g_2'(g_2^{-1}(u)) + f(b - g_2^{-1}(u))g_2''(g_2^{-1}(u))]}{[g_2'(g_2^{-1}(u))]^5}.$$

Using the fact that

$$f(x) = f(a) + (x-a)f'(a) + (x-a)^2 f''(a + \gamma_1(x-a)) = f(b) + (x-b)f'(b) + (x-b)^2 f''(b + \gamma_2(x-b)),$$
$$f'(x) = f'(a) + (x-a)f''(a + \theta_1(x-a)) = f'(b) + (x-b)f''(b + \theta_2(x-b))$$

for some $\gamma_1, \gamma_2, \theta_1, \theta_2 \in [0,1]$ and then combining (A.2), (A.3) and (A.4), I get

$$E\tilde{f}(x) = f(x)\int_{[x-b-g_2(b-a)]/h}^{[x-a+g_1(b-a)]/h} K(t)\,\mathrm{d}t - h\int_{[x-b-g_2(b-a)]/h}^{[x-a+g_1(b-a)]/h} tK(t)\,\mathrm{d}t + \frac{h^2}{2}\left[\int_{(x-b)/h}^{(x-a)/h} t^2K(t)f''(x - \lambda_{0t}ht)\,\mathrm{d}t + \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\mathcal{R}_1(x,t)\,\mathrm{d}t + \int_{[x-b-g_2(b-a)]/h}^{(x-b)/h} K(t)\mathcal{R}_2(x,t)\,\mathrm{d}t\right],$$

where

$$\mathcal{R}_1(x,t) = \left(t - \frac{x-a}{h}\right)^2 \mathcal{G}_1(\lambda_{1t}(ht - x + a)) + \frac{2(x-a)t}{h} f''(a + \theta_1(x-a)) - 2\left(\frac{x-a}{h}\right)^2 f''(a + \gamma_1(x-a)),$$
$$\mathcal{R}_2(x,t) = \left(t - \frac{x-b}{h}\right)^2 \mathcal{G}_2(\lambda_{2t}(-ht + x - b)) + \frac{2(x-b)t}{h} f''(b + \theta_2(x-b)) - 2\left(\frac{x-b}{h}\right)^2 f''(b + \gamma_2(x-b)).$$

By Assumption D, $h \to 0$ as $n \to \infty$, so when n is sufficiently large, $h \leqslant \min(g_1(b-a), g_2(b-a))$ then for any $x \in [a,b]$,

$$\int_{[x-b-g_2(b-a)]/h}^{[x-a+g_1(b-a)]/h} K(t)\,\mathrm{d}t = \int_{-1}^{1} K(t)\,\mathrm{d}t = 1, \quad \int_{[x-b-g_2(b-a)]/h}^{[x-a+g_1(b-a)]/h} tK(t)\,\mathrm{d}t = \int_{-1}^{1} tK(t) = 0,$$

since $x - a + g_1(b-a) \geqslant g_1(b-a) \geqslant h$ and $x - b - g_2(b-a) \leqslant -g_2(b-a) \leqslant -h$. Hence,

$$\begin{aligned}\left|E\tilde{f}(x) - f(x)\right| &= \frac{h^2}{2}\left|\int_{(x-b)/h}^{(x-a)/h} t^2K(t)f''(x - \lambda_{0t}ht)\,\mathrm{d}t + \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\mathcal{R}_1(x,t)\,\mathrm{d}t + \int_{[x-b-g_2(b-a)]/h}^{(x-b)/h} K(t)\mathcal{R}_2(x,t)\,\mathrm{d}t\right| \\ &\leqslant \frac{h^2}{2}\left|\int_{(x-b)/h}^{(x-a)/h} t^2K(t)f''(x - \lambda_{0t}ht)\,\mathrm{d}t\right| + \left|\frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\mathcal{R}_1(x,t)\,\mathrm{d}t\right| \\ &\quad + \left|\frac{h^2}{2}\int_{[x-b-g_2(b-a)]/h}^{(x-b)/h} K(t)\mathcal{R}_2(x,t)\,\mathrm{d}t\right|. \end{aligned} \tag{A.5}$$

Now I shall show the three terms on the right-hand side of inequality (A.5) are uniformly bounded for all $x \in [a,b]$.

First, by Assumption B,

$$\overline{f} \equiv \sup_{z\in[a,b]} f(z) < \infty, \quad \overline{f'} \equiv \sup_{z\in[a,b]} |f'(z)| < \infty, \quad \overline{f''} \equiv \sup_{z\in[a,b]} |f''(z)| < \infty. \tag{A.6}$$

Thus

$$\left|\int_{(x-b)/h}^{(x-a)/h} t^2K(t)f''(x-\lambda_{0t}ht)\,\mathrm{d}t\right| \leqslant \overline{f''}\int_{(x-b)/h}^{(x-a)/h} t^2K(t)\,\mathrm{d}t \leqslant \overline{f''}\int_{-\infty}^{\infty} t^2K(t)\,\mathrm{d}t = \kappa\overline{f''}. \tag{A.7}$$

Second, in the integral of (A.3), $\frac{x-a}{h} \leqslant t \leqslant \frac{x-a+g_1(b-a)}{h}$ and $0 \leqslant \lambda_{1t} \leqslant 1$, so $0 \leqslant \lambda_{1t}(ht-x+a) \leqslant g_1(b-a)$ and then $0 \leqslant g_1^{-1}(\lambda_{1t}(ht-x+a)) \leqslant b-a$ by the monotonicity of g_1^{-1}. Since $g_1(u) = u + d_1u^2 + A_1d_1^2u^3$, it is easy to see that

(i) $g_1'(u) \leqslant \max(1, 1+2d_1(b-a)+3A_1d_1^2(b-a)^2)$ and

$$g_1'(u) = 1 + 2d_1u + 3A_1d_1^2u^2 \geqslant \frac{3A_1-1}{3A_1} > 0 \tag{A.8}$$

for any $0 \leqslant u \leqslant b-a$;

(ii) $|g_1''(u)| = |2d_1 + 6A_1d_1^2u| \leqslant |2d_1| + 6A_1d_1^2(b-a)$ for any $0 \leqslant u \leqslant b-a$; and

(iii) $g_1'''(u) = 6A_1d_1^2 > 0$ is constant.

Define

$$\overline{g_1'} = \max(1, 1+2d_1(b-a)+3A_1d_1^2(b-a)^2), \quad \overline{g_1''} = |2d_1| + 6A_1d_1^2(b-a), \quad \overline{g_1'''} = 6A_1d_1^2,$$

then

$$\begin{aligned}\sup_{0\leqslant u\leqslant g_1(b-a)} |\mathcal{G}_1(u)| &\leqslant \sup_{0\leqslant u\leqslant b-a}\left|\frac{f''(a+u)g_1'(u)}{[g_1'(u)]^4}\right| + \sup_{0\leqslant u\leqslant b-a}\left|\frac{f(a+u)g_1'''(u)}{[g_1'(u)]^4}\right| \\ &\quad + 3\sup_{0\leqslant u\leqslant b-a}\left|\frac{f'(a+u)g_1'(u)g''(u)}{[g_1'(u)]^5}\right| + 3\sup_{0\leqslant u\leqslant b-a}\left|\frac{f(a+u)g_1''(u)^2}{[g_1'(u)]^5}\right| \\ &\leqslant \frac{\overline{f''}\cdot\overline{g_1'} + \overline{f}\cdot\overline{g_1'''}}{[1-1/(3A_1)]^4} + \frac{3\overline{f'}\cdot\overline{g_1'}\cdot\overline{g_1''} + 3\overline{f}\left(\overline{g_1''}\right)^2}{[1-1/(3A_1)]^5} \equiv \overline{\mathcal{G}_1} < \infty.\end{aligned}$$

Therefore,

$$\begin{aligned}&\left|\frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\mathcal{R}_1(x,t)\,\mathrm{d}t\right| = \left|\frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} t^2K(t)\mathcal{G}_1(\lambda_{1t}(ht-x+a))\,\mathrm{d}t\right. \\ &\qquad + \frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} \frac{2(x-a)}{h}tK(t)[f''(a+\theta_1(x-a)) - \mathcal{G}_1(\lambda_{1t}(ht-x+a))]\,\mathrm{d}t \\ &\qquad \left. + \frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} 2\left(\frac{x-a}{h}\right)^2 K(t)[\mathcal{G}_1(\lambda_{1t}(ht-x+a)) - f''(a+\gamma_1(x-a))]\,\mathrm{d}t\right| \\ &\leqslant \frac{h^2}{2}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} t^2K(t)\,|\mathcal{G}_1(\lambda_{1t}(ht-x+a))|\,\mathrm{d}t \\ &\qquad + h\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)|t|K(t)\,\left[\left|f''(a+\theta_1(x-a))\right| + |\mathcal{G}_1(\lambda_{1t}(ht-x+a))|\right]\,\mathrm{d}t \\ &\qquad + \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)^2K(t)\,\left[|\mathcal{G}_1(\lambda_{1t}(ht-x+a))| + \left|f''(a+\gamma_1(x-a))\right|\right]\,\mathrm{d}t \\ &\leqslant \frac{h^2}{2}\overline{\mathcal{G}_1}\int_{-\infty}^{\infty} t^2K(t)\,\mathrm{d}t + h(\overline{\mathcal{G}_1} + \overline{f''})\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)|t|K(t)\,\mathrm{d}t\end{aligned}$$

$$+ (\overline{\mathcal{G}_1} + \overline{f''}) \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)^2 K(t)\,\mathrm{d}t$$

$$= \frac{h^2}{2}\kappa\overline{\mathcal{G}_1} + h(\overline{\mathcal{G}_1} + \overline{f''}) \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)|t|K(t)\,\mathrm{d}t + (\overline{\mathcal{G}_1} + \overline{f''}) \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)^2 K(t)\,\mathrm{d}t$$

When $x > a + h$, $(x-a)/h > 1$ so $K(t) = 0$ on the entire interval of integration. Thus

$$\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)|t|K(t)\,\mathrm{d}t = \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)^2 K(t)\,\mathrm{d}t = 0.$$

When $a \leqslant x \leqslant a + h$,

$$\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)|t|K(t)\,\mathrm{d}t \leqslant h \int_{-\infty}^{\infty} |t|K(t)\,\mathrm{d}t \leqslant h \left[\int_{-\infty}^{\infty} |t|^2 K(t)\,\mathrm{d}t\right]^{1/2} = h\sqrt{\kappa},$$

where the last inequality follows from the fact that $K(t)$ is indeed a probability density function and the Liapounov's inequality, and,

$$\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} (x-a)^2 K(t)\,\mathrm{d}t \leqslant h^2 \int_{-\infty}^{\infty} K(t)\,\mathrm{d}t = h^2.$$

Combining these two cases, I have

$$\left|\frac{h^2}{2} \int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)\mathcal{R}_1(x,t)\,\mathrm{d}t\right| \leqslant h^2 \left[\frac{1}{2}\kappa\overline{\mathcal{G}_1} + \left(\overline{\mathcal{G}_1} + \overline{f''}\right)(1+\sqrt{\kappa})\right]. \tag{A.9}$$

Third, similarly, I have

$$\left|\frac{h^2}{2} \int_{[x-b-g_2(b-a)]/h}^{(x-b)/h} K(t)\mathcal{R}_2(x,t)\,\mathrm{d}t\right| \leqslant h^2 \left[\frac{1}{2}\kappa\overline{\mathcal{G}_2} + \left(\overline{\mathcal{G}_2} + \overline{f''}\right)(1+\sqrt{\kappa})\right]. \tag{A.10}$$

where

$$\overline{\mathcal{G}_2} \equiv \frac{\overline{f''} \cdot \overline{g_2'} + \overline{f} \cdot \overline{g_2'''}}{[1 - 1/(3A_2)]^4} + \frac{3\overline{f'} \cdot \overline{g_2'} \cdot \overline{g_2''} + 3\overline{f}\left(\overline{g_2''}\right)^2}{[1 - 1/(3A_2)]^5} < \infty$$

with

$$\overline{g_2'} = \max(1, 1 + 2d_2(b-a) + 3A_2 d_2^2 (b-a)^2), \quad \overline{g_2''} = |2d_2| + 6A_2 d_2^2 (b-a), \quad \overline{g_2'''} = 6A_2 d_2^2.$$

Finally, plugging (A.7), (A.9) and (A.10) into (A.5), I get

$$\left|E\tilde{f}(x) - f(x)\right| \leqslant h^2 \left[\frac{\kappa}{2}\left(\overline{f''} + \overline{\mathcal{G}_1} + \overline{\mathcal{G}_2}\right) + (1+\sqrt{\kappa})\left(2\overline{f''} + \overline{\mathcal{G}_1} + \overline{\mathcal{G}_2}\right)\right]. \tag{A.11}$$

Note that the coefficient of h^2 on the right-hand side doesn't depend on x (but depends on a and b) and hence the right-hand side is a uniform upper bound, so the desired result follows. □

A.2 Proof of Lemma 2

Since X_i, $i = 1, 2, \ldots, n$ are i.i.d. draws from f by Assumption A, then

$$\begin{aligned}
\operatorname{Var}\left(\tilde{f}(x)\right) &= \frac{1}{n}\operatorname{Var}\left\{\frac{I(a \leqslant X_1 \leqslant b)}{h}\left[K\left(\frac{x - X_1}{h}\right)\right.\right. \\
&\qquad \left.\left. + K\left(\frac{x - a + g_1(X_1 - a)}{h}\right) + K\left(\frac{b - x + g_2(b - X_1)}{h}\right)\right]\right\} \\
&\leqslant \frac{1}{n}E\left\{\frac{I(a \leqslant X_1 \leqslant b)}{h^2}\left[K\left(\frac{x - X_1}{h}\right)\right.\right. \\
&\qquad \left.\left. + K\left(\frac{x - a + g_1(X_1 - a)}{h}\right) + K\left(\frac{b - x + g_2(b - X_1)}{h}\right)\right]^2\right\} \\
&= \frac{1}{nh^2}\int_a^b K\left(\frac{x - z}{h}\right)^2 f(z)\,\mathrm{d}z + \frac{1}{nh^2}\int_a^b K\left(\frac{x - a + g_1(z - a)}{h}\right)^2 f(z)\,\mathrm{d}z \\
&\quad + \frac{1}{nh^2}\int_a^b K\left(\frac{b - x + g_2(b - z)}{h}\right)^2 f(z)\,\mathrm{d}z \\
&\quad + \frac{2}{nh^2}\int_a^b K\left(\frac{x - z}{h}\right) K\left(\frac{x - a + g_1(z - a)}{h}\right) f(z)\,\mathrm{d}z \\
&\quad + \frac{2}{nh^2}\int_a^b K\left(\frac{x - z}{h}\right) K\left(\frac{b - x + g_2(b - z)}{h}\right) f(z)\,\mathrm{d}z \\
&\quad + \frac{2}{nh^2}\int_a^b K\left(\frac{x - a + g_1(z - a)}{h}\right) K\left(\frac{b - x + g_2(b - z)}{h}\right) f(z)\,\mathrm{d}z, \qquad \text{(A.12)}
\end{aligned}$$

where the inequality follows from the fact that $\operatorname{Var}(Z) = E(Z^2) - (EZ)^2 \leqslant E(Z^2)$ for any random variable Z.

Let $R_K \equiv \int_{-1}^{1} K(u)^2\,\mathrm{d}u$. By Assumption C, K is bounded and has finite support, so $R_K < \infty$. In (A.12),

$$\frac{1}{nh^2}\int_a^b K\left(\frac{x - z}{h}\right)^2 f(z)\,\mathrm{d}z = \frac{1}{nh}\int_{(x-b)/h}^{(x-a)/h} K(t)^2 f(x - ht)\,\mathrm{d}t \leqslant \frac{\overline{f}}{nh}\int_{-\infty}^{\infty} K(t)^2\,\mathrm{d}t = \frac{R_K \overline{f}}{nh}. \qquad \text{(A.13)}$$

where $\overline{f} = \sup_{z \in [\underline{x}, \overline{x}]} f(z) < \infty$ as defined in (A.6). And

$$\frac{1}{nh^2}\int_a^b K\left(\frac{x - a + g_1(z - a)}{h}\right)^2 f(z)\,\mathrm{d}z = \frac{1}{nh}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)^2 \frac{f(a + g_1^{-1}(ht - x + a))}{g_1'(g_1^{-1}(ht - x + a))}\,\mathrm{d}t.$$

Since $\frac{x-a}{h} \leqslant t \leqslant \frac{x-a+g_1(b-a)}{h}$, so $0 \leqslant ht - x + a \leqslant g_1(b - a)$ and then $0 \leqslant g_1^{-1}(ht - x + a) \leqslant b - a$ by monotonicity of g_1^{-1}. Then, using (A.8), I have

$$\begin{aligned}
\frac{1}{nh^2}\int_a^b K\left(\frac{x - a + g_1(z - a)}{h}\right)^2 f(z)\,\mathrm{d}z &\leqslant \frac{1}{nh}\frac{\overline{f}}{1 - 1/(3A_1)}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)^2\,\mathrm{d}t \\
&\leqslant \frac{1}{nh}\frac{\overline{f}}{1 - 1/(3A_1)}\int_{-\infty}^{\infty} K(t)^2\,\mathrm{d}t = \frac{1}{nh}\frac{R_K \overline{f}}{1 - 1/(3A_1)}, \qquad \text{(A.14)}
\end{aligned}$$

Similarly,

$$\frac{1}{nh^2}\int_a^b K\left(\frac{b - x + g_2(b - z)}{h}\right)^2 f(z)\,\mathrm{d}z \leqslant \frac{1}{nh}\frac{R_K \overline{f}}{1 - 1/(3A_2)}. \qquad \text{(A.15)}$$

Let $\overline{K} \equiv \sup_{u\in[-1,1]} K(u)$ then by Assumption C I have $\overline{K} < \infty$. By similar changing variable procedure, I can get

$$\begin{aligned}
&\frac{2}{nh^2}\int_a^b K\left(\frac{x-z}{h}\right)K\left(\frac{x-a+g_1(z-a)}{h}\right)f(z)\,\mathrm{d}z \\
&\qquad = \frac{2}{nh}\int_{(x-b)/h}^{(x-a)/h} K(t)K\left(\frac{x-a+g_1(x-a-ht)}{h}\right)f(x-ht)\,\mathrm{d}t \\
&\qquad \leqslant \frac{2\overline{K}\cdot\overline{f}}{nh}\int_{-\infty}^{\infty} K(t)\,\mathrm{d}t = \frac{2\overline{K}\cdot\overline{f}}{nh}, \qquad (A.16)
\end{aligned}$$

and similarly,

$$\frac{2}{nh^2}\int_a^b K\left(\frac{x-z}{h}\right)K\left(\frac{b-x+g_2(b-z)}{h}\right)f(z)\,\mathrm{d}z \leqslant \frac{2\overline{K}\cdot\overline{f}}{nh}. \qquad (A.17)$$

By changing variable and applying similar argument to show (A.14), I can also get

$$\begin{aligned}
&\frac{2}{nh^2}\int_a^b K\left(\frac{x-a+g_1(z-a)}{h}\right)K\left(\frac{b-x+g_2(b-z)}{h}\right)f(z)\,\mathrm{d}z \\
&= \frac{2}{nh}\int_{(x-a)/h}^{[x-a+g_1(b-a)]/h} K(t)K\left(\frac{b-x+g_2(b-a-g_1^{-1}(ht-x+a))}{h}\right)\frac{f(a+g_1^{-1}(ht-x+a))}{g_1'(g_1^{-1}(ht-x+a))}\,\mathrm{d}t \\
&\leqslant \frac{2}{nh}\frac{\overline{K}\cdot\overline{f}}{1-1/(3A_1)}\int_{-\infty}^{\infty} K(t)\,\mathrm{d}t = \frac{2}{nh}\frac{\overline{K}\cdot\overline{f}}{1-1/(3A_1)}. \qquad (A.18)
\end{aligned}$$

Plugging (A.13), (A.14), (A.15), (A.16), (A.17) and (A.18) into (A.12), I finally get

$$\operatorname{Var}\left(\tilde{f}(x)\right) \leqslant \frac{1}{nh}\left[R_K\overline{f}\left(1+\frac{3A_1}{3A_1-1}+\frac{3A_2}{3A_2-1}\right)+2\overline{K}\cdot\overline{f}\frac{9A_1-2}{3A_1-1}\right]. \qquad (A.19)$$

Note that the coefficient of $(nh)^{-1}$ on the right-hand side doesn't depend on x (and even doesn't depend on a or b). Therefore the right-hand side is a uniform upper bound for the variance of $\tilde{f}(x)$ and then the desired result follows. □

A.3 Proof of Lemma 3

I prove (2.11) first. Using the c_r inequality,

$$\begin{aligned}
E\{[f_1^*(a+h_1)-f(a+h_1)]^2 \mid X_p\} &= E\left\{[f_1^*(a+h_1)-Ef_1^*(a+h_1)+Ef_1^*(a+h_1)-f(a+h_1)]^2 \mid X_p\right\} \\
&\leqslant 2E\left\{[f_1^*(a+h_1)-Ef_1^*(a+h_1)]^2 \mid X_p\right\}+2[Ef_1^*(a+h_1)-f(a+h_1)]^2. \qquad (A.20)
\end{aligned}$$

Denote $U_i = I(a \leqslant X_i \leqslant b)K\left(\frac{h_1-X_i+a}{h_1}\right)$, $i = 1,2,\ldots,n$. Since X_i's are i.i.d., then so are U_i's. By definition of $f_1^*(a+h_1)$,

$$\begin{aligned}
&E\{[f_1^*(a+h_1)-Ef_1^*(a+h_1)]^2 \mid X_p\} \\
&\quad = \frac{1}{n^2{h_1}^2}E\left\{\left[\sum_{i=1}^n I(a\leqslant X_i\leqslant b)K\left(\frac{h_1-X_i+a}{h_1}\right)-\sum_{i=1}^n E\left(I(a\leqslant X_i\leqslant b)K\left(\frac{h_1-X_i+a}{h_1}\right)\right)\right]^2 \,\middle|\, X_p\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2{h_1}^2} E\left\{\left[(U_p - EU_p) + \left(\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i - E\left(\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i\right)\right)\right]^2 \middle| X_p\right\} \\
&\leqslant \frac{2}{n^2{h_1}^2} E\left[(U_p - EU_p)^2 \mid X_p\right] + \frac{2}{n^2{h_1}^2} E\left\{\left[\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i - E\left(\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i\right)\right]^2 \middle| X_p\right\} \\
&= \frac{2}{n^2{h_1}^2}(U_p - EU_p)^2 + \frac{2}{n^2{h_1}^2} E\left[\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i - E\left(\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i\right)\right]^2 \\
&= \frac{2}{n^2{h_1}^2}(U_p - EU_p)^2 + \frac{2}{n^2{h_1}^2} \operatorname{Var}\left(\sum_{1\leqslant i\leqslant n,\, i\neq p} U_i\right) \\
&= \frac{2}{n^2{h_1}^2}(U_p - EU_p)^2 + \frac{2(n-1)}{n^2{h_1}^2} \operatorname{Var}(U_1) \\
&\leqslant \frac{2}{n^2{h_1}^2}(U_p - EU_p)^2 + \frac{2}{n{h_1}^2} EU_1^2, \qquad \text{(A.21)}
\end{aligned}
$$

where the first inequality follows from the c_r inequality and the second inequality is due to the fact that $\operatorname{Var}(U_1) = EU_1^2 - (EU_1)^2 \leqslant EU_1^2$. Since for all $i = 1, 2, \ldots, n$, $0 \leqslant U_i = I(a \leqslant X_i \leqslant b)K\left(\frac{h_1 - X_i + a}{h_1}\right) \leqslant \overline{K}$ where $\overline{K} = \sup_{u\in\mathbb{R}} K(u) < \infty$, then $0 \leqslant |EU_i| = EU_i \leqslant \overline{K}$ and

$$
\frac{1}{n^2{h_1}^2}(U_p - EU_p)^2 \leqslant \frac{(|U_p| + |EU_p|)^2}{n^2{h_1}^2} \leqslant \frac{(2\overline{K})^2}{n^2{h_1}^2} = O\left(\frac{1}{n^2{h_1}^2}\right).
$$

And because

$$
EU_1^2 = \int_a^b K\left(\frac{h_1 - z + a}{h_1}\right)^2 f(z)\,\mathrm{d}z = h_1 \int_{1-(b-a)/h_1}^1 K(t)^2 f(a + (1-t)h_1)\,\mathrm{d}t \leqslant h_1 \overline{f} \int_{-\infty}^{\infty} K(t)^2\,\mathrm{d}t,
$$

where $\overline{f} = \sup_{z\in[a,b]} f(z)$. By Assumption C, $R_K \equiv \int_{-\infty}^{\infty} K(t)^2\,\mathrm{d}t < \infty$ and hence $EU_1^2/(n{h_1}^2) \leqslant h_1 R_K \overline{f}/(n{h_1}^2) = O\left(\frac{1}{nh_1}\right)$. Provided that $nh_1 \to \infty$ as $n \to \infty$, I get

$$
E\left\{[f_1^*(a+h_1) - Ef_1^*(a+h_1)]^2 \mid X_p\right\} \leqslant O\left(\frac{1}{n^2{h_1}^2}\right) + O\left(\frac{1}{nh_1}\right) = O\left(\frac{1}{nh_1}\right). \qquad \text{(A.22)}
$$

Also, X_i's are i.i.d. by Assumption A, so I have

$$
\begin{aligned}
Ef_1^*(a+h_1) &= \frac{1}{h_1} E\left[I(a \leqslant X_1 \leqslant b)K\left(\frac{h_1 - X_1 + a}{h_1}\right)\right] \\
&= \frac{1}{h_1}\int_a^b K\left(\frac{h_1 - z + a}{h_1}\right) f(z)\,\mathrm{d}z = \int_{1-(b-a)/h_1}^1 K(t) f(a + h_1 - th_1)\,\mathrm{d}t \\
&= f(a+h_1)\int_{1-(b-a)/h_1}^1 K(t)\,\mathrm{d}t - h_1 f'(a+h_1)\int_{1-(b-a)/h_1}^1 tK(t)\,\mathrm{d}t \\
&\qquad + {h_1}^2 \int_{1-(b-a)/h_1}^1 t^2 K(t) f''(a + h - \lambda t h_1)\,\mathrm{d}t,
\end{aligned}
$$

where $\lambda \in [0, 1]$. Since $h_1 \to 0$ as $n \to \infty$, so $h_1 < (b-a)/2$ when n is sufficiently large, then using

$\int_{1-(b-a)/h_1}^{1} K(t)\,\mathrm{d}t = \int_{-1}^{1} K(t)\,\mathrm{d}t = 1$ and $\int_{1-(b-a)/h_1}^{1} tK(t)\,\mathrm{d}t = \int_{-1}^{1} tK(t)\,\mathrm{d}t = 0$, I have

$$Ef_1^*(a+h_1) - f(a+h_1) = {h_1}^2 \int_{-1}^{1} t^2 K(t) f''(a+h-\lambda t h_1)\,\mathrm{d}t \leqslant {h_1}^2 \overline{f''} \int_{-1}^{1} t^2 K(t)\,\mathrm{d}t = {h_1}^2 \overline{f''}\kappa = O({h_1}^2),$$

where $\overline{f''} = \sup_{z\in[a,b]} f''(z)$. Hence,

$$[Ef_1^*(a+h_1) - f(a+h_1)]^2 = O({h_1}^4). \tag{A.23}$$

Now (2.11) is proven by combining (A.20), (A.22) and (A.23).

By letting $U_i = I(a \leqslant X_i \leqslant b)K\left(\frac{h_1+X_i-b}{h_1}\right)$, I can show (2.13) with analogous argument. To show (2.12) and (2.14), I first can use similar argument to show

$$E\left\{[f_1^*(a) - f(a)]^2 \mid X_p\right\} = O\left({h_0}^4 + \frac{1}{nh_0}\right), \quad E\left\{[f_2^*(b) - f(b)]^2 \mid X_p\right\} = O\left({h_0}^4 + \frac{1}{nh_0}\right)$$

Then the desired results follow from (2.6). □

A.4 Proof of Lemma 4

I only show (2.15) and (2.16) here. (2.17) and (2.18) can be proven analogously.

By definition of $f_1(a+h_1)$, using the c_r inequality, I have

$$\begin{aligned} E\left\{[f_1(a+h_1) - f(a+h_1)]^2 \mid X_p\right\} &= E\left\{\left[f_1^*(a+h_1) - f(a+h_1) + \frac{1}{n^2}\right]^2 \middle| X_p\right\} \\ &\leqslant 2E\left\{[f_1^*(a+h_1) - f(a+h_1)]^2 \mid X_p\right\} + 2E\left[\left(\frac{1}{n^2}\right)^2 \middle| X_p\right] \\ &= O\left({h_1}^4 + \frac{1}{nh_1}\right) + \frac{2}{n^4}. \end{aligned}$$

Recall that I have known the fastest convergence rate of the first term in the right-hand side is of $O(n^{-4/5})$, which is slower than the second term, whence (2.15) follows.

To see (2.16), by definition of $f_1(a)$,

$$E\left\{[f_1(a) - f(a)]^2 \mid X_p\right\} = E\left\{\left[\max\left(f_1^*(a), 1/n^2\right) - f(a)\right]^2 \middle| X_p\right\}.$$

Because $f_1^*(a) \leqslant \max\left(f_1^*(a), 1/n^2\right) \leqslant f_1^*(a) + 1/n^2$,

$$\left|\max\left(f_1^*(a), \frac{1}{n^2}\right) - f(a)\right| \leqslant |f_1^*(a) - f(a)| + \left|f_1^*(a) + \frac{1}{n^2} - f(a)\right|,$$

and then I have

$$E\left\{[f_1(a) - f(a)]^2 \mid X_p\right\} \leqslant E\left\{\left[|f_1^*(a) - f(a)| + \left|f_1^*(a) - f(a) + \frac{1}{n^2}\right|\right]^2 \middle| X_p\right\}$$

$$\leqslant 2E\left\{[f_1^*(a)-f(a)]^2 \mid X_p\right\}+2E\left\{\left[f_1^*(a)-f(a)+\frac{1}{n^2}\right]^2 \middle| X_p\right\}$$
$$\leqslant 2E\left\{[f_1^*(a)-f(a)]^2 \mid X_p\right\}+4E\left\{[f_1^*(a)-f(a)]^2 \mid X_p\right\}+\frac{4}{n^4}$$
$$=O\left({h_1}^4+\frac{1}{nh_1}\right)+O\left(\frac{1}{n^4}\right).$$

where the second and the third inequalities follow from the c_r inequality. Then by the same argument to show (2.15), I have (2.16) proven. □

A.5 Proof of Lemma 5

I only show the conclusion for $E[(\hat{d}_1-d_1)^2 \mid X_p]$ here. By repeatedly applying the c_r inequality, I can get

$$\begin{aligned}
E\left[(\hat{d}_1-d_1)^2 \middle| X_p\right] &= E\left\{\left[\left(\hat{d}_1-\frac{\log f(a+h_1)-\log f(a)}{h_1}\right)+\left(\frac{\log f(a+h_1)-\log f(a)}{h_1}-d\right)\right]^2 \middle| X_p\right\} \\
&\leqslant 2E\left[\left(\hat{d}_1-\frac{\log f(a+h_1)-\log f(a)}{h_1}\right)^2 \middle| X_p\right] \\
&\qquad +2E\left[\left(\frac{\log f(a+h_1)-\log f(a)}{h_1}-d\right)^2 \middle| X_p\right] \\
&= 2E\left[\left(\frac{\log f_1(a+h_1)-\log f(a+h_1)}{h_1}-\frac{\log f_1(a)-\log f(a)}{h_1}\right)^2 \middle| X_p\right] \\
&\qquad +2\left(\frac{\log f(a+h_1)-\log f(a)}{h_1}-d\right)^2 \\
&\leqslant \frac{4}{{h_1}^2}E\left\{[\log f_1(a+h_1)-\log f(a+h_1)]^2 \mid X_p\right\} \\
&\qquad +\frac{4}{{h_1}^2}E\left\{[\log f_1(a)-\log f(a)]^2 \mid X_p\right\}+2\left(\frac{\log f(a+h_1)-\log f(a)}{h_1}-d\right)^2 \\
&= \frac{4}{{h_1}^2}J_1+\frac{4}{{h_1}^2}J_2+2J_3,
\end{aligned} \tag{A.24}$$

where

$$J_1=E\left\{[\log f_1(a+h_1)-\log f(a+h_1)]^2 \mid X_p\right\},$$
$$J_2=E\left\{[\log f_1(a)-\log f(a)]^2 \mid X_p\right\},$$
$$J_3=\left(\frac{\log f(a+h_1)-\log f(a)}{h_1}-d\right)^2.$$

First, by the mean value theorem, I have

$$J_1=E\left\{\left[\frac{f_1(a+h_1)-f(a+h_1)}{\lambda f_1(a+h_1)+(1-\lambda)f(a+h_1)}\right]^2 \middle| X_p\right\}$$
$$\leqslant E\left\{\left[\frac{f_1(a+h_1)-f(a+h_1)}{(1-\lambda)f(a+h_1)}\right]^2 \middle| X_p\right\}=\frac{E\left\{[f_1(a+h_1)-f(a+h_1)]^2 \mid X_p\right\}}{[(1-\lambda)f(a+h_1)]^2}$$

$$
\begin{aligned}
&\leqslant \frac{1}{(1-\lambda)^2 c_0^2} E\left\{[f_1(a+h_1) - f(a+h_1)]^2 \mid X_p\right\} \\
&= O\left({h_1}^4 + \frac{1}{nh_1}\right),
\end{aligned} \tag{A.25}
$$

where $\lambda \in (0,1)$ is a constant, the first inequality is because the estimator $f_1(a+h_1) > 0$ by definition, the second inequality is because $f(a+h_1) \geqslant c_0 > 0$ by Assumption B, and the last equality is obtained by using Lemma 4.

Similarly, by using the fact that $f_1(a) > 0$ and $f(a) \geqslant c_0 > 0$, I have

$$
J_2 \leqslant \frac{1}{(1-\lambda')^2 c_0^2} E\left\{[f_1(a) - f(a)]^2 \mid X_p\right\} = O\left({h_1}^4 + \frac{1}{nh_1}\right), \tag{A.26}
$$

where $\lambda' \in (0,1)$ is a constant.

For J_3, by second-order Taylor expansion of $\log f(\cdot)$, I have

$$
\begin{aligned}
J_3 &= \left\{ \frac{1}{h_1}\left[h_1 \frac{f'(a)}{f(a)} + \frac{{h_1}^2}{2} \cdot \frac{f''(a+\lambda'' h_1) f(a+\lambda'' h_1) - [f'(a+\lambda'' h_1)]^2}{[f(a+\lambda'' h_1)]^2} \right] - \frac{f'(a)}{f(a)} \right\}^2 \\
&= \frac{{h_1}^2}{4} \left\{ \frac{f''(a+\lambda'' h_1) f(a+\lambda'' h_1) - [f'(a+\lambda'' h_1)]^2}{[f(a+\lambda'' h_1)]^2} \right\}^2 \\
&\leqslant \frac{{h_1}^2}{4} \cdot \frac{\left[\overline{f''} \cdot \overline{f} + (\overline{f'})^2\right]^2}{c_0^4} \\
&= O({h_1}^2),
\end{aligned} \tag{A.27}
$$

where $\lambda'' \in [0,1]$ is a constant. Here $\overline{f''} = \sup_{z \in [a,b]} |f''(z)| < \infty$, $\overline{f'} = \sup_{z \in [a,b]} |f'(z)| < \infty$ and $\overline{f} = \sup_{z \in [a,b]} f(z) < \infty$. Then combining (A.24), (A.25), (A.26) and (A.27), I get

$$
E\left[(\hat{d}_1 - d_1)^2 \,\middle|\, X_p \right] = O\left({h_1}^2 + \frac{1}{n{h_1}^3}\right) + O\left({h_1}^2 + \frac{1}{n{h_1}^3}\right) + O({h_1}^2) = O\left({h_1}^2 + \frac{1}{n{h_1}^3}\right).
$$

The other conclusion can be proven analogously. □

A.6 Proof of Theorem 2

By triangular inequality,

$$
\sup_{x \in [a,b]} \left| \hat{f}(x) - f(x) \right| = \sup_{x \in [a,b]} \left| \hat{f}(x) - \tilde{f}(x) + \tilde{f}(x) - f(x) \right| \leqslant \sup_{x \in [a,b]} \left| \hat{f}(x) - \tilde{f}(x) \right| + \sup_{x \in [a,b]} \left| \tilde{f}(x) - f(x) \right|.
$$

Since I have already know that the second term above is of order $O\left(h^2 + \sqrt{\log n/(nh)}\right)$ almost surely and hence $O_p\left(h^2 + \sqrt{\log n/(nh)}\right)$ by Theorem 1, I am left with the first term.

By definition of $\hat{f}$ and $\tilde{f}$, I have for any $x \in [a,b]$,

$$
\left| \hat{f}(x) - \tilde{f}(x) \right| = \left| \frac{1}{nh} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant b) \left[K\left(\frac{x - a + \hat{g}_1(X_i - a)}{h} \right) - K\left(\frac{x - a + g_1(X_i - a)}{h} \right) \right.\right.
$$

$$
\begin{aligned}
&\quad + K\left(\frac{b-x+\hat{g}_2(b-X_i)}{h}\right) - K\left(\frac{b-x+g_2(b-X_i)}{h}\right)\Bigg]\Bigg| \\
&\leqslant \frac{1}{nh}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant b)\left[K\left(\frac{x-a+\hat{g}_1(X_i-a)}{h}\right) - K\left(\frac{x-a+g_1(X_i-a)}{h}\right)\right]\right| \\
&\qquad + \frac{1}{nh}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant b)\left[K\left(\frac{b-x+\hat{g}_2(b-X_i)}{h}\right) - K\left(\frac{b-x+g_2(b-X_i)}{h}\right)\right]\right| \\
&= J_1 + J_2,
\end{aligned}
\tag{A.28}
$$

where

$$
\begin{aligned}
J_1 &= \frac{1}{nh}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant b)\left[K\left(\frac{x-a+\hat{g}_1(X_i-a)}{h}\right) - K\left(\frac{x-a+g_1(X_i-a)}{h}\right)\right]\right|, \\
J_2 &= \frac{1}{nh}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant b)\left[K\left(\frac{b-x+\hat{g}_2(b-X_i)}{h}\right) - K\left(\frac{b-x+g_2(b-X_i)}{h}\right)\right]\right|.
\end{aligned}
$$

Consider J_1 first. Because for any $d \in \mathbb{R}$, $u \geqslant 0$ and $A > 1/3$,

$$
u + du^2 + Ad^2u^3 = (1 + du + Ad^2u^2)u = \left[\left(\sqrt{A}du + \frac{1}{2\sqrt{A}}\right)^2 + \frac{4A-1}{4A}\right]u \geqslant \frac{4A-1}{4A}u,
$$

it is easy to see that $g_1(u) > h$, $\hat{g}_1(u) > h$ for $u > \rho h$, where $\rho = 4A_1/(4A_1 - 1)$. Hence, for $x \in [a, b]$, when $X_i - a > \rho h$,

$$
K\left(\frac{x-a+\hat{g}_1(X_i-a)}{h}\right) = K\left(\frac{x-a+g_1(X_i-a)}{h}\right) = 0,
$$

since $K(u) = 0$ for $|u| > 1$ by Assumption C. By Assumption D, $h \to 0$ as $n \to \infty$, thus by applying first-order Taylor expansion, I have for sufficiently large n such that $h < (b-a)/\rho$,

$$
\begin{aligned}
J_1 &= \frac{1}{nh}\sum_{i=1}^{n}\Bigg|I(a \leqslant X_i \leqslant a+\rho h)\Bigg[\frac{\hat{g}_1(X_i-a) - g_1(X_i-a)}{h} \\
&\qquad\qquad \times K'\left(\frac{x-a+\lambda_i\hat{g}_1(X_i-a) + (1-\lambda_i)g_1(X_i-a)}{h}\right)\Bigg]\Bigg| \\
&\leqslant \frac{\overline{K'}}{nh^2}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant a+\rho h)\left[\hat{g}_1(X_i-a) - g_1(X_i-a)\right]\right| \\
&= \frac{\overline{K'}}{nh^2}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant a+\rho h)\left[(\hat{d}_1 - d_1)(X_i-a)^2 + A_1(\hat{d}_1^2 - d_1^2)(X_i-a)^3\right]\right| \\
&\leqslant \frac{\overline{K'}}{nh^2}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant a+\rho h)(X_i-a)^2(\hat{d}_1 - d_1)\right| \\
&\qquad\qquad + \frac{A_1\overline{K'}}{nh^2}\sum_{i=1}^{n}\left|I(a \leqslant X_i \leqslant a+\rho h)(X_i-a)^3(\hat{d}_1^2 - d_1^2)\right|, \\
&\leqslant \frac{\overline{K'}}{nh^2}\sum_{i=1}^{n}\rho^2h^2\left|I(a \leqslant X_i \leqslant a+\rho h)(\hat{d}_1 - d_1)\right| + \frac{A_1\overline{K'}}{nh^2}\sum_{i=1}^{n}\rho^3h^3\left|I(a \leqslant X_i \leqslant a+\rho h)(\hat{d}_1^2 - d_1^2)\right|,
\end{aligned}
$$

where $\lambda_i \in [0, 1]$ and $\overline{K'} = \sup_{u\in\mathbb{R}}|K'(u)| < \infty$. The last inequality is obtained by using the fact that $0 \leqslant X_i - a \leqslant \rho h$ when $a \leqslant X_i \leqslant a + \rho h$. Because $|\hat{d}_1^2 - d_1^2| = |(\hat{d}_1 - d_1)^2 + 2d_1(\hat{d}_1 - d_1)| \leqslant (\hat{d}_1 - d_1)^2 +$

$2|d_1||\hat{d}_1 - d_1|$, then by triangular inequality, I have

$$J_1 \leqslant \left(\rho^2 h^2 \overline{K'} + 2A_1\rho^3 h^3 |d_1| \overline{K'}\right) \cdot \frac{1}{nh^2} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)|\hat{d}_1 - d_1| + 2A_1\rho^3 h^4 |d_1| \overline{K'} \cdot \frac{1}{nh^3} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)(\hat{d}_1 - d_1)^2. \quad \text{(A.29)}$$

Under Assumptions D and E, $h_1 \to 0$ and $n{h_1}^3 \to \infty$ as $n \to \infty$, and Lemma 5 gives $E[(\hat{d}_1 - d_1)^2 \mid X_i] = O\left({h_1}^2 + \frac{1}{n{h_1}^3}\right) = O\left(\max\left({h_1}^2, \frac{1}{n{h_1}^3}\right)\right)$ for all $i = 1, 2, \ldots, n$. Thus by the Hölder's inequality,

$$E\left[|\hat{d}_1 - d_1| \,\middle|\, X_i\right] \leqslant \left\{E\left[(\hat{d}_1 - d_1)^2 \mid X_i\right]\right\}^{1/2} = O\left(\max\left({h_1}^2, \frac{1}{n{h_1}^3}\right)\right)^{1/2} = O\left(\max\left(h_1, \frac{1}{\sqrt{n{h_1}^3}}\right)\right).$$

It means that for sufficiently large n,

$$E\left[|\hat{d}_1 - d_1| \,\middle|\, X_i\right] \leqslant C_1 \max\left(h_1, \frac{1}{\sqrt{n{h_1}^3}}\right), \quad E\left[(\hat{d}_1 - d_1)^2 \,\middle|\, X_i\right] \leqslant C_2 \max\left({h_1}^2, \frac{1}{n{h_1}^3}\right),$$

for some constants $C_1, C_2 > 0$. Then, for any $M > 0$, by Markov's inequality,

$$\begin{aligned}
\Pr\left(\frac{1}{nh^2}\sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)|\hat{d}_1 - d_1| \geqslant M\right) &\leqslant \frac{1}{M} E\left[\frac{1}{nh^2}\sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)|\hat{d}_1 - d_1|\right] \\
&= \frac{1}{nh^2 M}\sum_{i=1}^{n} E\left[I(a \leqslant X_i \leqslant a + \rho h) E\left(|\hat{d}_1 - d_1| \mid X_i\right)\right] \\
&\leqslant \frac{1}{nh^2 M}\sum_{i=1}^{n} C_1 \max\left(h_1, \frac{1}{\sqrt{n{h_1}^3}}\right) \cdot E[I(a \leqslant X_i \leqslant a + \rho h)] \\
&= \frac{C_1}{h^2 M} \max\left(h_1, \frac{1}{\sqrt{n{h_1}^3}}\right) \int_a^{a+\rho h} f(z)\, \mathrm{d}z \\
&\leqslant \frac{C_1}{h^2 M} \max\left(h_1, \frac{1}{\sqrt{n{h_1}^3}}\right) \overline{f} \rho h = \frac{C_1 \overline{f} \rho}{M} \cdot \frac{\max\left(h_1, \frac{1}{\sqrt{n{h_1}^3}}\right)}{h}
\end{aligned}$$

and similarly

$$\begin{aligned}
\Pr\left(\frac{1}{nh^3}\sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)(\hat{d}_1 - d_1)^2 \geqslant M\right) &\leqslant \frac{1}{M} E\left[\frac{1}{nh^3}\sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)(\hat{d}_1 - d_1)^2\right] \\
&= \frac{1}{nh^3 M}\sum_{i=1}^{n} E\left\{I(a \leqslant X_i \leqslant a + \rho h) E\left[(\hat{d}_1 - d_1)^2 \mid X_i\right]\right\} \\
&\leqslant \frac{1}{nh^3 M}\sum_{i=1}^{n} C_2 \max\left({h_1}^2, \frac{1}{n{h_1}^3}\right) \cdot E[I(a \leqslant X_i \leqslant a + \rho h)]
\end{aligned}$$

$$\leqslant \frac{C_2 \overline{f} \rho}{M} \cdot \frac{\max\left(h_1{}^2, \frac{1}{nh_1{}^3}\right)}{h^2},$$

where $\overline{f} = \sup_{z\in[a,b]} f(z)$. Since $h_1 = O(h)$ and $1/\sqrt{nh_1{}^3} = O(h)$, then there exists a constant $C_3 > 0$ such that for sufficiently large n,

$$\frac{\max\left(h_1, \frac{1}{\sqrt{nh_1{}^3}}\right)}{h} \leqslant C_3, \quad \frac{\max\left(h_1{}^2, \frac{1}{nh_1{}^3}\right)}{h^2} \leqslant C_3,$$

which implies that for any $\epsilon > 0$, I can pick M sufficiently large, e.g. $M = C_3 \overline{f} \rho \max(C_1, C_2)/\epsilon$, such that

$$\Pr\left(\frac{1}{nh^2} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)|\hat{d}_1 - d_1| \geqslant M\right) \leqslant \epsilon,$$
$$\Pr\left(\frac{1}{nh^3} \sum_{i=1}^{n} I(a \leqslant X_i \leqslant a + \rho h)(\hat{d}_1 - d_1)^2 \geqslant M\right) \leqslant \epsilon.$$

Therefore, it follows from (A.29) that

$$J_1 \leqslant (\rho^2 h^2 \overline{K'} + 2A_1 \rho^3 h^3 |d_1| \overline{K'}) \cdot O_p(1) + 2A_1 \rho^3 h^4 |d_1| \overline{K'} \cdot O_p(1) = O_p(h^2) + O_p(h^3) + O_p(h^4) = O_p(h^2). \tag{A.30}$$

Note that the upper bound for J_1 given in (A.29) doesn't depend on x, so the order of J_1 given in (A.30) is uniform for all $x \in [a, b]$. Similarly, I have $J_2 = O_p(h^2)$ uniformly for all $x \in [a, b]$. Hence,

$$\sup_{x\in[a,b]} \left|\hat{f}(x) - \tilde{f}(x)\right| \leqslant \sup_{x\in[a,b]} J_1 + \sup_{x\in[a,b]} J_2 = O_p(h^2),$$

which completes the proof of the theorem. □