Journal of Nonparametric Statistics
Vol. 00, No. 00, Month 2010, 1–35

RESEARCH ARTICLE

Large Sample Results for Varying Kernel Regression Estimates

Hira L. Koul & Weixing Song
Michigan State University and Kansas State University

(*June 24, 2013*)

The varying kernel density estimates are particularly designed for positive random variables. Unlike the commonly used symmetric kernel density estimates, the varying kernel density estimates do not suffer from the boundary problem. This paper establishes asymptotic normality and uniform almost sure convergence results for a varying kernel density estimate when the underlying random variable is positive. Similar results are also obtained for a varying kernel nonparametric estimate of the regression function when the covariate is positive. Pros and cons of the varying kernel regression estimate are also discussed via a simulation study.

Keywords: Varying Kernel Regression; Inverse Gamma Distribution; Almost Sure Convergence; Central Limit Theorem.

AMS Subject Classification: 62G08; 62G20

Research supported in part by the NSF DMS Collaborative Grants 1205271 and 1205276.
*Corresponding author: Weixing Song, Email: weixing@ksu.edu

ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

1. Introduction

In this paper we investigate consistency and asymptotic normality of a varying kernel density estimator when the density function is supported on $(0, \infty)$. We also propose a varying kernel regression function estimator when covariate in the underlying regression model is non-negative and investigate its similar asymptotic properties.

The problem of estimating density function of a random variable X taking values on the real line has been of a long-lasting research interest among statisticians, and numerous interesting and fundamental results have been obtained. Kernel density estimation method no doubt is the most popular among all the proposed nonparametric procedures. In the commonly used kernel estimation set up, the kernel function K is often chosen to be a density function symmetric around 0 satisfying some moment conditions.

With a random sample $X_1, X_2, \ldots, X_n$ of X, and a bandwidth h depending on n, the kernel estimate of density function f of X is $\hat{f}(x) = (nh)^{-1} \sum_{i=1}^{n} K((x - X_i)/h)$. The contribution from each sample point X_i to $\hat{f}(x)$ is mainly controlled by how far X_i is from x on the h-scale. Moreover, because of the symmetry of K, the sample points on either sides of x with the same distance from x make the same contribution to the estimate. Consequently, symmetric kernel assigns positive weights outside the density support set near the boundaries, which is also the very reason why the commonly used symmetric kernel density estimates have the unpleasant boundary problem. This boundary problem is also present in the Nadaraya-Watson (N-W) estimators of a nonparametric regression function. Numerous ways have been proposed to remove the boundary effect. In the context of density estimation, see Schuster (1985), Marron and Ruppert (1994), Jones (1993), Fan and Gijbels (1992), Cowling and Hall (1996) etc.; for nonparametric regression, see Gasser and Müller (1979), Müller (1991), Müller and Wang (1994), John (1984) and the references therein.

The research on estimating the density functions not supported on the entire real line using asymmetric density kernels started from late 1990s. When density has a compact support, motivated by the Bernstein polynomial approximation theorem in mathematical function analysis, Chen (1999) proposed Beta kernel density estimators and analyzed bias and variance of these estimators. By reversing the role of estimation point and data point in Chen (1999)'s estimation procedure, and using the Gaussian copula kernel, Jones and Henderson (2007) proposed two density estimators. When density functions are supported on $(0, \infty)$, Chen (2000b) constructed a Gamma kernel density estimate and Scaillet (2004) proposed an inverse Gaussian kernel and a reciprocal inverse Gaussian kernel density estimate. Chaubey et al. (2012) also proposed a density estimator for non-negative random variables via smoothing of the empirical distribution function using a generalization of Hille's lemma. A varying kernel density estimate, which is an asymmetric kernel density estimate and based on a modification of Chen's Gamma kernel density estimate, was recently proposed by Mnatsakanov and Sarkisian (2012) (M-S).

Compared to the traditional symmetric kernel estimation procedures, there are two unique features about all of the above asymmetric kernel methods: (1) the smoothness of the density estimate is controlled by the shape or scale parameter of the asymmetric kernel and the location where the estimation is made; and (2) the asymmetric kernels have the same support as the density functions to be estimated, thus the kernels do not allocate any weight outside the support. As a consequence, all of the above asymmetric kernel density estimators can effectively reduce the boundary bias and they all achieve the optimal rate of convergence for the mean integrated squared error.

Some asymmetric density estimators are bona fide densities, such as the ones proposed by Jones and Henderson (2007). Most of them are not, but they become one after a slight modification, for example, the M-S varying kernel density estimate. In principle, the commonly used symmetric kernel density estimate, in which the kernel is supported over a symmetric interval around 0, can still be used for estimating the density function of a random variable with some restricted range, but the resulting estimate itself may not be a density function any more. For example, using standard normal kernel to estimate the density function of a positive random variable, the resulting kernel density estimate does not integrate to 1 over $(0, \infty)$.

Most of the research on asymmetric kernel estimation methodology has been focused on the density estimation, and the asymptotic theories are limited to the bias, variance, or mean square error derivations. To the best of our knowledge, literature is scant on the investigation of the consistency of asymmetric kernel density estimators except for Bouezmarni and Rolin (2003) and Chaubey et al. (2012). Nothing is available on their asymptotic distributions. The situation in the context of nonparametric regression is also surprising. Using Beta or Gamma kernel function, Chen (2000a, 2002) proposed the local linear estimators for regression function and derived their asymptotic bias and variance, but did not analyze their asymptotic distributions.

The present paper makes an attempt at filling this void by investigating the large sample properties of the M-S kernel procedure in the fields of both density and regression function estimations. First, in the context of density estimation, we investigate the asymptotic normality and uniform almost sure convergence of the M-S kernel density estimate. Secondly, in the context of nonparametric regression, we investigate the asymptotic behavior of the M-S kernel regression function estimate. We derive its asymptotic conditional bias and conditional variance, and establish its uniform almost sure consistency and the asymptotic normality. Thirdly, bandwidth selection is explored for the sake of implementing the methodology. As a byproduct, the paper provides a theoretical framework for investigating the similar properties of other asymmetric kernel estimators.

2. M-S Kernel Regression Estimation

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a population X supported on $(0, \infty)$. Let

$$K_\alpha^*(x,t) = \frac{1}{t\Gamma(\alpha)} \Big(\frac{\alpha x}{t}\Big)^\alpha \exp\Big(-\frac{\alpha x}{t}\Big), \quad \alpha > 0,\, t > 0,\, x > 0. \tag{2.1}$$

For a sequence of positive real numbers α_n, M-S proposed the following nonparametric estimate for the density of X:

$$f_{\alpha_n}^*(x) = \frac{1}{n}\sum_{i=1}^n K_{\alpha_n}^*(x, X_i) = \frac{1}{n}\sum_{i=1}^n \frac{1}{X_i\Gamma(\alpha_n)}\Big(\frac{\alpha_n x}{X_i}\Big)^{\alpha_n} \exp\Big(-\frac{\alpha_n x}{X_i}\Big). \tag{2.2}$$

The estimate (2.2) is constructed using the technique of recovering a function from its Mellin transform applied in the moment-identifiable problem. There is a close connection between $K_\alpha^*(x,t)$ and the Gamma and inverse Gamma density functions. For each fixed t, $K_\alpha^*(\cdot, t)$ is a Gamma density function with scale parameter t/α and shape parameter $\alpha+1$; for each fixed x, $K_\alpha^*(x, \cdot)$ is the density function of an Inverse Gamma distribution with scale parameter α and shape parameter αx. Unfortunately, as seen in M-S, the asymptotic bias of $f_{\alpha_n}^*(x)$ depends on the first derivative of the underlying density function of X, which is due to the fact that $\alpha x/(\alpha - 1)$, instead of x, is the mean of the density $K_\alpha^*(x,t)$ viewed as a function of t. To reduce the bias, M-S used a modified version of $K_\alpha^*(x,t)$, viz,

$$K_\alpha(x,t) = \frac{1}{t\Gamma(\alpha+1)}\Big(\frac{\alpha x}{t}\Big)^{\alpha+1}\exp\Big(-\frac{\alpha x}{t}\Big), \tag{2.3}$$

to construct the density estimate. For fixed x, $K_\alpha(x,\cdot)$ now is the density function of an Inverse Gamma distribution with shape parameter $\alpha+1$ and scale parameter αx, the mean of which is exactly x; for fixed t, $K_\alpha(\cdot,t)$ is not a Gamma density function anymore, but $\alpha K_\alpha(x,t)/(\alpha+1)$ is a Gamma density function with shape parameter $\alpha+2$ and scale parameter t/α. These properties indeed imply a very interesting connection of the M-S kernel $K_\alpha(x,t)$ and the normal kernel used in the commonly used density estimate for large values of α. In fact, for a fixed x, let T_α be a random variable having density function $K_\alpha(x,\cdot)$, and for a fixed t, let X_α be a random variable having density function $\alpha K_\alpha(\cdot,t)/(\alpha+1)$. Then one can verify that

$$\sqrt{\alpha}\Big(T_\alpha/x-1\Big) \to_d N(0,1), \quad \sqrt{\alpha}\Big(X_\alpha/t-1\Big) \to_d N(0,1), \quad \text{as } \alpha\to\infty.$$

Here, and in the following, $\to_d$ denotes the convergence in distribution. If we let $h = 1/\sqrt{\alpha}$, then from the above facts it follows that as $\alpha\to\infty$,

$$K_\alpha(x,t) \approx \frac{1}{h}\phi\Big(\frac{x/t-1}{h}\Big) \quad \text{or} \quad K_\alpha(x,t) \approx \frac{1}{h}\phi\Big(\frac{t/x-1}{h}\Big),$$

where $\phi(\cdot)$ denotes the standard normal density function. Therefore, the M-S kernel K_α approximately behaves like the standard normal kernel, while the distance between x and t is not the usual Euclidean distance $|x-t|$, but rather the relative distance $|x-t|/t$ or $|x-t|/x$; for the commonly used kernel function, x and t are symmetric in the sense of difference, while in the kernel function $K_\alpha(x,t)$, x and t are asymptotically symmetric in the sense of division; the parameter $1/\sqrt{\alpha}$ plays the role of bandwidth as in commonly used kernel set up. To have a better understanding about the smoothing effect of the kernel function $K_\alpha(x,t)$, we plot the functions for a pseudo-data set $0.5, 1, 2, 3$ and $\alpha = 5, 20$ over the range $x \in (0,7)$ which are shown in Figure 1. Clearly, all the curves are skewed to the right implying that more weights are put on the values to the right of the observed data points; as α gets larger, all the curves shrink towards the data points and the shape of the kernels changes according to the values of the data points.

For a random sample $X_1, X_2, \ldots, X_n$ from the population X supported on $(0,\infty)$, the

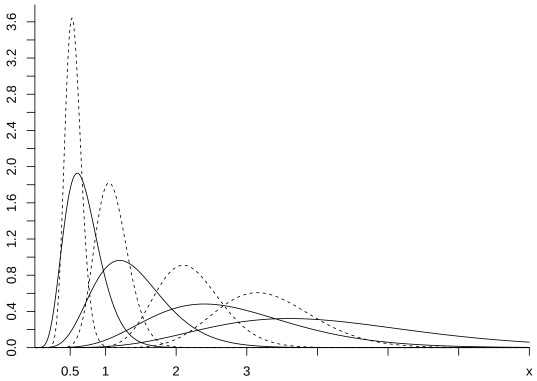

Figure 1. The kernel function $K_\alpha(x,t)$ for the 4 pseudo-data points listed in the text, and two choices of α. The solid curves are for $\alpha = 5$, and the dotted curves for $\alpha = 20$. The Curve with the highest peak is for the data point 0.5, with the second highest is for the data point 1, and so on.

M-S kernel density estimate based the modified kernel $K_\alpha(x,t)$ is

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\alpha_n}(x, X_i). \tag{2.4}$$

The expression for the mean squares error (MSE) of $\hat{f}_n(x)$ is derived in M-S, as well as the L_1-consistency. Different from the commonly used symmetric kernel density estimates, the M-S kernel density estimates do not suffer from the boundary effect, which is confirmed both by the theories developed and simulation studies conducted in M-S. Although $f(x)$ is not defined at $x = 0$, it is clear that $\hat{f}_n(0) = 0$ almost surely. This intrinsic constraint is only desirable if $\lim_{x\to 0} f(x) = 0$. Some other asymmetric kernel estimates also suffer from the similar disturbance, such as the inverse and reciprocal inverse Gaussian kernel estimates proposed in Scaillet (2004), the copula-based kernel estimate suggested by Jones and Henderson (2007). If $\lim_{x\to 0} f(x) > 0$, then to analyze the boundary behavior of $\hat{f}_n(x)$ around 0, similar to the symmetric kernel case, we analyze the limiting behavior of the bias in $\hat{f}_n(x)$ at $x = u/\alpha_n$, where $0 < u < 1$. This is done in Section 6.

There is no discussion in the literature on the asymptotic normality of the M-S kernel density estimate (2.4). This paper will try to fill this void, not just because this topic itself is very interesting, but also because it has some very practical implications, for example, knowing of the asymptotic distribution of $\hat{f}_n(x)$ enables us to construct confidence interval for the density function $f(x)$. Parallel to the commonly used symmetric kernel estimation methodology, and also as a further development, we also investigate the large sample behavior of the nonparametric estimators of regression function using the M-S kernel, when the covariate is positive.

The relationship between a scalar response Y and a covariate X is often investigated through the regression model $Y = m(X) + \varepsilon$, where ε is the random error and X is one dimensional and a positive random variable. Furthermore, we assume that $E(\varepsilon|X = x) = 0$ and $\sigma^2(x) := E(\varepsilon^2|X = x) > 0$, for almost all $x > 0$. Let $\{(X_i, Y_i), i = 1, 2, \ldots, n\}$ be a random sample from this regression model. Inspired by the construction of the N-W kernel regression estimate, the M-S kernel regression estimate of $m(x)$ is defined to be

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n K_{\alpha_n}(x, X_i)Y_i}{\sum_{i=1}^n K_{\alpha_n}(x, X_i)}, \tag{2.5}$$

In spite of the similarity between this estimate and its N-W kernel counterpart, the very different characteristics of the M-S kernel function from the commonly used symmetric kernel imply that many technical challenges encountered in the development of asymptotic theories for the new estimates are different. Under some regularity conditions on the underlying density function $f(x)$ and the regression function $m(x)$, asymptotic normality of the M-S kernel estimate $\hat{m}_n(x)$, as well as its uniform consistency, is established in the paper.

From the definition of K_{α_n}, one can derive a much simpler expression for $\hat{m}_n(x)$. In fact, after some cancelation, we have

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n X_i^{-\alpha_n-2}\exp(-\alpha_n x/X_i)Y_i}{\sum_{i=1}^n X_i^{-\alpha_n-2}\exp(-\alpha_n x/X_i)}.$$

This formula is mainly useful for the computation of $\hat{m}_n$ while (2.5) is convenient for theoretical development.

Being an asymmetric kernel, $K_\alpha^*(x, t)$ defined in (2.1), and $K_\alpha(x, t)$ defined in (2.3), are rather different from the asymmetric kernels discussed in Cline (1988) and Abadir and Lawford (2004). In (2.1) and (2.3), at each $x > 0$, the data points X_i's behave like the scale parameter of x, while in Cline (1988), Abadir and Lawford (2004), X_i's appear as the location parameter of x. So, the inadmissibility of the asymmetric kernel proved in Cline (1988) does not apply to the varying kernels defined in (2.1) and (2.3).

The proposed estimation procedure is mainly developed for a univariate X. It is desirable to seek its extensions for higher dimensional positive covariates. Similar to the commonly used symmetric kernel regression, one way to proceed is to use the product kernel in the definition of the regression function estimate. Another way is to use a multivariate extension of Gamma or inverse Gamma density function as the kernel function. The product kernel method is the most straightforward and natural choice, and similar theoretical results as in 1-dimension can be easily derived. However, using multivariate extensions of Gamma or inverse Gamma density as the kernel may not be practical, since the multivariate Gamma density functions proposed in the literature all have complicated forms, which makes the computation and theoretical development of the corresponding varying kernel estimates much challenging. For some definitions of multivariate Gamma distribution, see Kotz et al. (2000).

The paper is organized as follows. Section 3 discusses the large sample results about $\hat{m}_n(x)$ along with the needed technical assumptions. In particular, it contains an approximate expression for the conditional MSE, and a central limit theorem, and a uniform consistency result about $\hat{m}_n$. Section 4 contains a discussion on the selection of the smoothing parameter α_n. Findings of a simulation study are presented in Section 5, and the proofs of the main results appear in Section 6. Unless specified otherwise, all limits are taken as $n \to \infty$.

3. Large Sample Results of $\hat{m}_n(x)$

We start with analyzing the asymptotic properties of the conditional bias and conditional variance, hence the conditional MSE, of $\hat{m}_n(x)$ defined in (2.5). Then a typical application of Lindeberg-Feller central limit theorem will lead to the asymptotic normality of $\hat{m}_n(x)$. As a byproduct, the asymptotic normality of the M-S kernel estimate $\hat{f}_n(x)$ is also a natural consequence. Thus, confidence intervals for the true density function and regression function can be constructed. Finally, uniform almost sure convergence results of $\hat{f}_n(x)$ and $\hat{m}_n(x)$ over any bounded sub-intervals of $(0, \infty)$ are developed by using the Borel-Cantelli lemma after verifying the Cramér condition for the M-S kernel function.

The following is a list of technical assumptions used for deriving these results.

(A1). The second order derivatives of $f(x)$ is continuous and bounded on $(0, \infty)$.

(A2). The second order derivatives of $f(x)m(x)$ is continuous and bounded on $(0, \infty)$.

(A3). The second order derivative of $\sigma^2(x) = E(\varepsilon^2|X = x)$ is continuous and bounded for all $x > 0$.

(A4). For some $\delta > 0$, the second order derivative of $E(|\varepsilon|^{2+\delta}|X = x)$ is continuous and bounded in $x \in (0, \infty)$.

(A5). $\alpha_n \to \infty$, $\sqrt{\alpha_n}/n \to 0$.

Condition (A1) on $f(x)$ is the same as the one adopted by M-S when deriving the bias and variance of $\hat{f}_n(x)$. Condition (A3) is required for dealing with the large sample argument pertaining to the random error, and is not needed if one is willing to assume the homoscedasticity. Condition (A4) is needed in proving the asymptotic normality of the proposed estimators, while (A5) is a minimal condition needed for the smoothing parameter. Additional assumptions on α_n as needed are stated in various theorems presented below.

In the following, for any function $g(x)$, $g'(x)$, $g''(x)$ denote the first and second derivatives of $g(x)$, respectively.

3.1. *Bias and Variance*

The following theorem presents the asymptotic expansions of the conditional biases and the variances, hence the conditional MSE, of $\hat{m}_n(x)$. Let

$$b(x) := x^2 \left[\frac{m'(x)f'(x)}{f(x)} + \frac{m''(x)}{2} \right], \qquad v(x) := \frac{\sigma^2(x)}{2xf(x)\sqrt{\pi}}, \tag{3.1}$$

and $\mathbf{X} := \{X_1, X_2, \dots, X_n\}$.

Theorem 3.1 : *Suppose the assumptions (A1), (A2), (A3), and (A5) hold. Then, for any $x \in (0, \infty)$ with $f(x) > 0$,*

$$\text{Bias}(\hat{m}_n(x)|\mathbf{X}) = \frac{b(x)}{\alpha_n} + o_p\Big(\frac{1}{\alpha_n}\Big) + O_p\left(\frac{1}{\sqrt{n\sqrt{\alpha_n}}}\right), \tag{3.2}$$

$$\text{Var}(\hat{m}_n(x)|\mathbf{X}) = \frac{v(x)\sqrt{\alpha_n}}{n} + o_p\Big(\frac{\sqrt{\alpha_n}}{n}\Big). \tag{3.3}$$

Thus the conditional MSE of $\hat{m}_n(x)$ has the asymptotic expansion

$$MSE(\hat{m}_n(x)|\mathbf{X}) = \frac{b^2(x)}{\alpha_n^2} + \frac{v(x)\sqrt{\alpha_n}}{n} + o_p\Big(\frac{1}{\alpha_n^2}\Big) + o_p\Big(\frac{\sqrt{\alpha_n}}{n}\Big) + o_p\left(\frac{1}{\sqrt{n}\alpha_n^{5/4}}\right).$$

Remark: The unconditional version of Theorem 3.1 is very hard to derive. This is also true for N-W kernel regression. Although Härdle et al. (2004) indicated that the conditional MSE of N-W kernel regression estimate could be derived from a linearization technique, and the result is summarized in Theorem 4.1 in Härdle et al. (2004), the rigorous proof is not provided. But we can show that the unconditional version of Theorem 3.1 remains valid for

$$\hat{m}_n^*(x) = \frac{\sum_{i=1}^n K_{\alpha_n}(x, X_i)Y_i}{n^{-2} + \sum_{i=1}^n K_{\alpha_n}(x, X_i)},$$

a slightly modified version of $\hat{m}_n(x)$. The similar idea was used in Fan (1993) when dealing with the local linear regression, and a proof of the unconditional MSE of $\hat{m}_n^*(x)$ can follow the same thread as the proof of Theorem 3 in Fan (1993).

Recalling the above discussion on the analogy between α_n and the bandwidth in the commonly used symmetric kernel density estimate, one can easily see the similarity of the bias and variance expressions between the M-S kernel estimate and the N-W kernel estimate.

Similar to the N-W kernel regression case, one can choose the optimal smoothing parameter $\alpha_{n,\text{opt}}$ by minimizing the leading term in the conditional MSE of $\hat{m}_n$ with respect to α_n. We can verify that $\alpha_{n,\text{opt}}$ has the order of $n^{2/5}$, with the corresponding MSE having the order of $n^{-4/5}$. Recall the same order is obtained for the N-W kernel regression estimate based on the same criterion.

3.2. *Asymptotic Normality*

First, we give the asymptotic normality of the M-S kernel density estimate.

Theorem 3.2: *Suppose the assumptions (A1), (A4), and (A5) hold. Then for any* $x \in (0, \infty)$ *with* $f(x) > 0$,

$$\Big(\frac{f(x)\sqrt{\alpha_n}}{2xn\sqrt{\pi}}\Big)^{-1/2}\Big[\hat{f}_n(x) - f(x) - \frac{x^2 f''(x)}{2(\alpha_n - 1)}\Big] \to_d N(0, 1).$$

The asymptotic normality of $\hat{f}_n(x)$ implies that $\hat{f}_n(x)$ converges to $f(x)$ in probability, hence $1/\hat{f}_n(x)$ converges to $1/f(x)$ in probability, whenever $f(x) > 0$. This result is used in the proof of the asymptotic normality of $\hat{m}_n(x)$, which is stated in the next theorem.

Theorem 3.3: *Suppose the assumptions in Theorem 3.1 hold. Then, for any $x \in (0,\infty)$ with $f(x) > 0$,*

$$\Big(\frac{v(x)\sqrt{\alpha_n}}{n}\Big)^{-1/2}\Big[\hat{m}_n(x) - m(x) - \frac{b(x)}{\alpha_n - 1}\Big] \to_d N(0,1),$$

where $b(x)$ and $v(x)$ are defined in (3.1).

It is noted that there is a non-negligible asymptotic bias appearing in the above results, a characteristic shared with the N-W kernel regression estimate. This bias can be eliminated by under-smoothing which, in the current set up, is to select a larger α_n such that $\sqrt{n}/\alpha_n^{5/4} \to 0$ without violating conditions $\alpha_n \to \infty$, $\sqrt{\alpha_n}/n \to 0$. The large sample confidence intervals for $m(x)$ thus can be constructed with the help of Theorem 3.3.

3.3. *Almost Sure Uniform Convergence*

In this section we develop an almost sure uniform convergence result for $\hat{m}_n(x)$ over an arbitrary bounded sub-interval of $(0,\infty)$. In the N-W kernel regression estimation scenario, a similar result is obtained by using the Borel-Cantelli lemma and the Bernstein inequality, but the Cramér condition must be verified before applying these well known results. That is, For any fixed $x > 0$, $k \geq 2$, we have to show that

$$E|K_\alpha(x,X)|^k \leq k!\Big(\frac{c\sqrt{\alpha}}{n}\Big)^{k-2} EK_\alpha^2(x,X)$$

for some positive constant c when α is large.

The following two theorems give the almost sure uniform convergence of $\hat{f}_n$ to f and $\hat{m}_n$ to m over bounded sub-intervals of $(0,\infty)$.

Theorem 3.4: *In addition to (A1) and (A5), assume that $\alpha_n^{1/2}\log n/n \to 0$. Then for any constants a and b such that $0 < a < b < \infty$,*

$$\sup_{x\in[a,b]}\Big|\hat{f}_n(x) - f(x)\Big| = O\Big(\frac{1}{\alpha_n}\Big) + o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big), \quad a.s.$$

Theorem 3.5: *In addition to (A1) to (A5), assume that $\alpha_n^{1/2}\log n/n \to 0$. Then for* *any constants a and b such that $0 < a < b < \infty$,*

$$\sup_{x\in[a,b]} \left|\hat{m}_n(x) - m(x)\right| = O\Big(\frac{1}{\alpha_n}\Big) + o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big), \quad a.s.$$

By assuming some stronger conditions on the tails of f and m at the boundaries, the above uniform almost sure convergence results can be extended to be over some suitable intervals increasing to $(0, \infty)$. However, we do not pursue it here simply because of the involved technical details and lack of a useful application.

4. Selection of Smoothing Parameters

It is well known that the smoothing parameter plays a crucial role in nonparametric kernel regression. Abundant research has been conducted for the N-W kernel type regression estimation methodology, see, e.g., Wand and Jones (1994) and Hart (1997) for more data-driven choices of the smoothing parameters in this setup. However, to the best of our knowledge, there is no work done for the asymmetric kernel regression.

In this section we propose several smoothing parameter selection procedures for implementing the M-S kernel technique. First, we recall the least square cross-validation (LSCV) procedure from M-S, and discuss its extension, k-fold LSCV. Secondly, we propose the smoothing parameter selection procedures in the nonparametric regression setup. The k-fold LSCV and the generalized cross-validation (GCV) will be discussed. These procedures are analogous to the commonly used data-driven procedures used in the N-W kernel regression estimation context. The theoretical properties, such as the consistency of these smoothing parameter selectors to some "optimal" smoothing parameter, might be discussed in the similar way as in John (1984), Härdle et al. (1992, 1988) and the references therein. However, we will not investigate this important topic in the current paper, which deserves an independent in-depth study.

4.1. *Density Estimation: k-fold LSCV*

The motivation of the LSCV comes from expanding the MISE of $\hat{f}$. Define

$$LSCV(\alpha) = \int \hat{f}^2(x)dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i),$$

where $\hat{f}_{-i}(X_i)$ is the leave-one-out M-S kernel density estimate for $f(X_i)$ without using the i-th observation. Then the LSCV smoothing parameter is defined by $\hat{\alpha}_{\text{LSCV}} = \text{argmin}_{\alpha} LSCV(\alpha)$. For the M-S kernel density estimate (2.4),

$$LSCV(\alpha) = \frac{\Gamma(2\alpha+3)}{n^2\alpha\Gamma^2(\alpha+1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{(X_iX_j)^{\alpha+1}}{(X_i+X_j)^{2\alpha+3}} - \frac{2}{n(n-1)\Gamma(\alpha+1)}\sum_{i\neq j}\frac{1}{X_j}\Big(\frac{\alpha X_i}{X_j}\Big)^{\alpha+1}\exp\Big(-\frac{\alpha X_i}{X_j}\Big).$$

An extension of the above leave-one-out LSCV is the k-fold LSCV procedure. First split the data into k roughly equal-sized parts; then for each part, calculate the prediction error based on the M-S kernel density estimate constructed from all data from other $k-1$ parts; finally, take the sum of the k prediction errors as the quantity to be minimized. In particular, for our current setup, the k-fold $LSCV$ has a similar structure as the leave-one-out LSCV except for the second term now defined as

$$\frac{2}{n\Gamma(\alpha+1)}\sum_{i=1}^{n}\Big[\frac{1}{n-n_i}\sum_{j\notin D(i)}\frac{1}{X_j}\Big(\frac{\alpha X_i}{X_j}\Big)^{\alpha+1}\exp\Big(-\frac{\alpha X_i}{X_j}\Big)\Big],$$

where $D(i)$ is the set of indices of the data part including X_i. For convenience, if we use $D_1, D_2, \ldots, D_k$ to denote the data subscripts in the first part, second part and so on, then $D(i) = \{j : i, j \in D_l, l = 1, 2, \ldots, k\}$, and n_i is the size of $D(i)$. The k-fold LSCV will reduce to the leave-one-out LSCV when $k = n$.

4.2. *M-S Kernel Regression: k-fold LSCV*

The basic idea of LSCV in regression setup is to select the smoothing parameter by minimizing prediction error. For this purpose, let $\hat{m}_{D/D(i)}(X_i)$ be the M-S kernel estimate of $m(x)$ at $x = X_i$ of the same type as $\hat{m}_n(x)$ except that it is computed without using the data parts including the i-th observation (X_i, Y_i), where $D = \{1, 2, \ldots, n\}$. The LSCV smoothing parameter $\hat{\alpha}_{\text{LSCV}}$ is the value of α that minimizes the LSCV criterion

$$\begin{aligned}\text{CV}(\alpha) &= \sum_{i=1}^{n}[Y_i - \hat{m}_{D/D(i)}(X_i)]^2 \\ &= \sum_{i=1}^{n}\Big[Y_i - \frac{\sum_{j\notin D(i)}^{n} X_j^{-\alpha-2}\exp(-\alpha X_i/X_j)Y_j}{\sum_{j\notin D(i)}^{n} X_j^{-\alpha-2}\exp(-\alpha X_i/X_j)}\Big]^2.\end{aligned}$$

The independence between (X_i, Y_i) and $\hat{m}_{D/D(i)}(X_i)$ indicates that $CV(\alpha)$ will give an accurate assessment of how well the estimate $\hat{m}_n(x)$ will predict future observations.

4.3. *M-S Kernel Regression: Generalized Cross-validation (GCV)*

The GCV procedure from the N-W kernel regression can also be adapted to the current setup. Define

$$w_{ij} = \frac{X_j^{-\alpha-2}\exp(-\alpha X_i/X_j)}{\sum_{k=1}^n X_k^{-\alpha-2}\exp(-\alpha X_i/X_k)}, \quad i,j = 1,2,\ldots,n.$$

Then the GCV smoothing parameter $\hat{\alpha}_{\mathrm{GCV}}$ is the value of α that minimizes the GCV criterion $GCV(\alpha)$ defined as

$$\mathrm{GCV}(\alpha) = \frac{n\sum_{i=1}^n\left[Y_i - \sum_{j=1}^n w_{ij}Y_j\right]^2}{\left[n - \sum_{i=1}^n w_{ii}\right]^2}.$$

There is no one smoothing parameter selection procedure which is uniformly superior to others, in the sense that the selected smoothing values always produce estimates with smallest MSE. The simulation study conducted in the next section shows that for some data sets, a selection procedure might not even work. A common practice is to try several procedures, and make an overall evaluation to decide a proper smoothing value.

5. Simulation Study

To evaluate the finite sample performance of the proposed M-S kernel regression estimates, we conducted a simulation study. In the simulation, the underlying density function of the design variable is chosen to be log-normal with $\mu = 0, \sigma = 1$, and the random error ε to be normal with mean 0 and standard deviation 0.5. Two simple regression functions, $m(x) = 1/x^2$, $m(x) = (x-1.5)^2$, are considered. For $m(x) = 1/x^2$, the estimate will be evaluated at 1024 equally spaced values over the interval $(0.1, 1)$; for $m(x) = (x-1.5)^2$, the estimate will be evaluated at 1024 equally spaced values over the interval $(0, 3)$, and the sample sizes used are 100 and 200. Then the MSEs between the estimated values and true values of the regression function will be used for comparison.

It is always controversial when comparing two different nonparametric smoothing pro-

Table 1. MSE comparison: $m(x) = 1/x^2, x \in (0.1, 1)$

	M-S kernel			N-W Kernel		
n	LSCV	GCV	$n^{2/5}$	LSCV	GCV	$h = n^{-1/5}$
100	(22) 7.944	(12) 54.522	119.113	×	×	(0.398) 148.185
200	(29) 5.116	(10) 10.530	14.497	×	(0.009) 2.359	(0.347) 174.172

cedures, especially when one or both procedures involve the smoothing parameters which play a crucial role in determine the smoothness of the fitted regression function, since by selecting a proper smoothing parameter, one method can often be made to outperform the other. Therefore, for the sake of fairness in comparison, one should use the same criterion to select the smoothing parameters whenever possible. Unfortunately, sometimes the chosen criterion works for one procedures, but does not work for another. In this case, one might try different criteria for both procedures and make an overall comparison. The 5-fold LSCV and GCV criteria are tried to select the bandwidth for both M-S kernel and N-W kernel estimate. The standard normal kernel is used to construct the N-W kernel estimate.

Table 1 presents the simulation study when $m(x) = 1/x^2$. The numbers within the parentheses are the smoothing values selected by various criteria, and the numbers outside the parentheses are the MSEs. For $n = 100$, the 5-fold LSCV criterion and GCV do not work for the N-W procedure and a crossed sign is used in the table to indicate this case. For $n = 200$, LSCV still does not work for the N-W estimator, while GCV works. Also, $h = n^{-1/5}$, the bandwidth based on the optimal order of the conditional MSE, is used to calculate the N-W estimate. The 5-fold LSCV criterion works for M-S kernel estimate. We also try $\alpha = n^{2/5}$, the smoothing α value based on the optimal order of the conditional MSE for M-S kernel estimate, to calculate the M-S kernel estimate. The values in the parentheses are the values of smoothing parameters. Figure 2 provides a visual comparison between these two procedures. To keep the figure neat, we only plot the M-S kernel estimate with LSCV bandwidth, and the N-W kernel estimate with $h = n^{-1/5}$. Here, and in the subsequent figures, the thick solid curve denotes the true regression function, the thin solid line denotes the M-S kernel estimate, and the dashed line is for the N-W estimate. Clearly, with respect to the boundary area, the M-S kernel estimate does better than the N-W kernel estimate. We also tried GCV criterion to choose the smoothing parameters. Figure 3 provides a visual comparisons between M-S and N-W

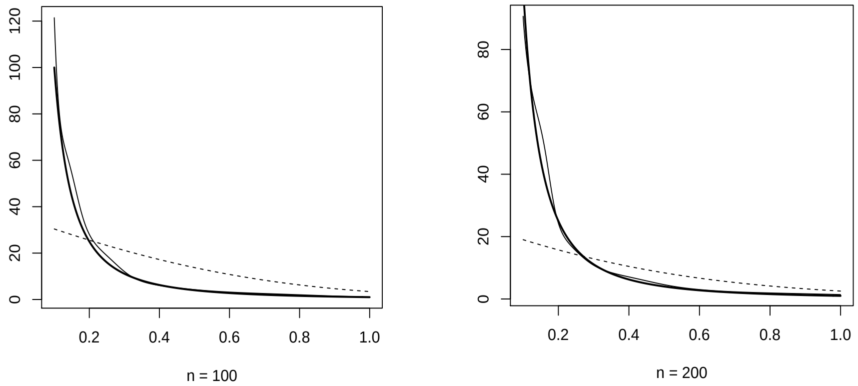

Figure 2. Estimates of $m(x) = 1/x^2$. Smoothing values are selected by LSCV and optimal order of MSE.

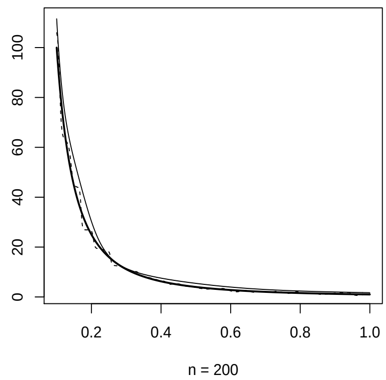

Figure 3. Estimates of $m(x) = 1/x^2$. Smoothing values are selected by GCV.

procedures with smoothing values selected by GCV. The MSEs reported in Table 1 and Figure 3 clearly indicate that the GCV favors the N-W kernel estimate more than the M-S kernel estimate, although the N-W kernel estimate possesses a larger variability.

We also tried fitting the regression function using the boundary kernel suggested by Gasser and Müller (1979). The MSEs are generally smaller than the N-W kernel estimates, but still much larger than the M-S kernel estimates. For example, for $h = n^{-1/5}$, the MSEs using the boundary kernel are 162.133 when $n = 100$, and 38.62 when $n = 200$.

Table 2 reports the MSEs from the simulation study when $m(x) = (x - 1.5)^2$. Now the 5-fold LSCV and GCV criteria work for both procedures. The M-S kernel estimate with both LSCV and GCV bandwidths outperforms the N-W kernel estimates with all the

Table 2. MSE comparison: $m(x) = (x - 1.5)^2, x \in (0, 3)$

	M-S kernel			N-W Kernel		
n	LSCV	GCV	$n^{2/5}$	LSCV	GCV	$h = n^{-1/5}$
100	(71)0.025	(30)0.016	(6.310)0.049	(0.429)0.086	(0.089)0.065	(0.398)0.030
200	(54)0.014	(37)0.012	(8.326)0.055	(0.917)0.250	(0.090)0.067	(0.347)0.030

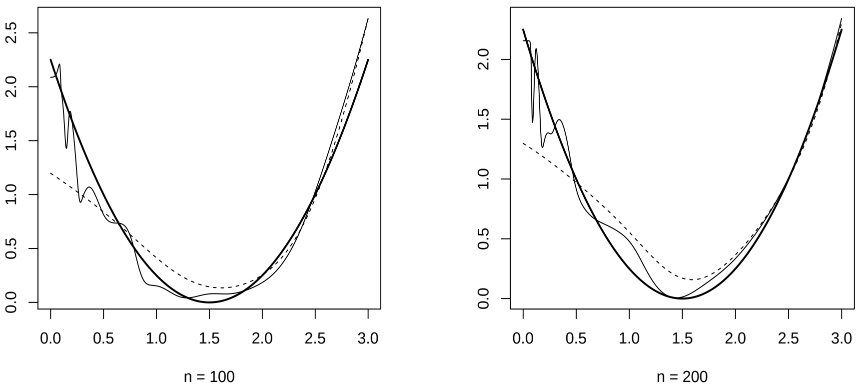

Figure 4. Estimates of $m(x) = (x - 1.5)^2$. Smoothing values are selected by LSCV and optimal order of MSE.

selected bandwidths, but the contrary is true when both procedures use the bandwidth by minimizing the AIMSE expression.

Figure 4 shows the fitted curves from the the M-S kernel estimate with LSCV bandwidth, and the N-W kernel estimate with $h = n^{-1/5}$. It is clear that the M-S kernel estimator is a very promising competitor for N-W kernel estimator. This point is reconfirmed by Figure 5, which shows the fitted curves from both kernel estimators with smoothing values selected by by GCV criterion.

6. Proofs of Main Results

This section contains the proofs of all the large sample results presented in Section 2. Inverse Gamma density function and its moments will be repeatedly referred to in the following proofs. For convenience, we list all the needed results here. Density function of

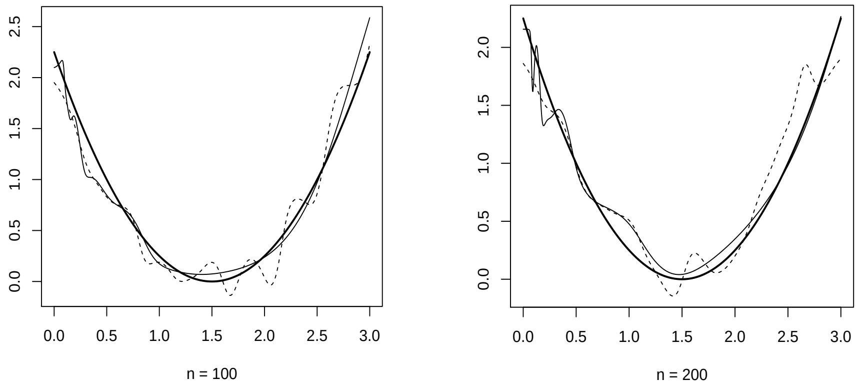

Figure 5. Estimates of $m(x) = (x - 1.5)^2$. Smoothing values are selected by GCV.

an inverse Gamma distribution with shape parameter p and rate parameter λ is

$$g(u, p, \lambda) = \frac{\lambda^p}{\Gamma(p)} \Big(\frac{1}{u}\Big)^{p+1} \exp\Big(-\frac{\lambda}{u}\Big), \quad u > 0.$$

Its mean μ, variance τ^2, and the fourth central moment ν_4, respectively, are

$$\mu = \frac{\lambda}{p-1}, \quad \tau^2 = \frac{\lambda^2}{(p-1)^2(p-2)},$$
$$\nu_4 = \frac{\lambda^4(3p+15)}{(p-1)^4(p-2)(p-3)(p-4)}.$$

Let

$$p_k = k(\alpha_n + 2) - 1, \quad \lambda_k = k\alpha_n x, \quad k = 1, 2, \cdots, \ x > 0.$$

Write μ_k, τ_k, and ν_{4k} for μ, τ, and ν_4 when λ and p are replaced by λ_k and p_k, respectively. The following lemma on the inverse Gamma distribution is crucial for the subsequent arguments.

Lemma 6.1: *Let $l(u)$ be a function such that the second order derivative of $l(u)$ is continuous and bounded on $(0, \infty)$. Then, for α_n large enough, and for all $x > 0$ and $k \geq 1$,*

$$\int_0^\infty g(u; p_k, \lambda_k) l(u) du = l(x) + \frac{(2-2k)xl'(x)}{p_k - 1}$$
$$+ \frac{[(2-2k)^2(p_k - 2) + k^2\alpha_n^2]x^2 l''(x)}{2(p_k-1)^2(p_k-2)} + o\Big(\frac{1}{\alpha_n}\Big).$$

Proof of Lemma 6.1.

Fix an $x > 0$. Note that $\mu_k := \lambda_k/(p_k - 1) = x + (2-2k)x/(p_k-1)$. Taylor expansion of $l(\mu_k)$ around x up to the second order yields

$$l(\mu_k) = l(x) + \frac{(2-2k)xl'(x)}{p_k - 1} + \frac{(2-2k)^2x^2l''(\xi)}{2(p_k-1)^2}, \tag{6.1}$$

where ξ is some value between $x + (2-2k)x/(p_k - 1)$ and x. Recall μ_k is the mean of $g(u; p_k, \lambda_k)$. Taylor expansion of $l(u)$ around μ_k yields

$$\begin{aligned}&\int_0^\infty l(u)g(u, p_k, \lambda_k)du \qquad (6.2)\\ &= l(\mu_k) + \frac{1}{2}l''(\mu_k)\int_0^\infty (u-\mu_k)^2 g(u; p_k, \lambda_k)du\\ &\quad + \frac{1}{2}\int_0^\infty (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du\end{aligned}$$

for some $\tilde{u}$ between u and μ_k. From (6.1) and the continuity of l'', we can verify that the two leading terms on the right hand side of (6.2) match the expansion in the lemma. So it is sufficient to show that the third term on the right hand side of (6.2) is of the order $o(1/\alpha_n)$.

Since l'' is continuous, so it is uniformly continuous over any closed sub-intervals in $(0, \infty)$. For any $\epsilon > 0$, select a $0 < \gamma < x$, such that for any y with $|y - x| \le \gamma$, $|l''(x) - l''(y)| < \epsilon$.

Let $\delta_1 = x - \gamma/2$. The boundedness of l'' implies

$$\left|\int_0^{\delta_1} (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du\right| \le c\int_0^{\delta_1} (u-\mu_k)^2 g(u; p_k, \lambda_k)du.$$

Note that the inverse Gamma density function $g(u, p_k, \lambda_k)$ is unimodal, and the mode is $\alpha_n x/(\alpha_n + 2)$, which approaches x when $\alpha_n \to \infty$. Therefore, for α_n large enough, $\delta_1 < \alpha_n x/(\alpha_n + 2)$, and for all $u \in (0, \delta_1)$, $g(u, p_k, \lambda_k) \le g(\delta_1, p_k, \lambda_k)$. Hence

$$\int_0^{\delta_1} (u-\mu_k)^2 g(u; p_k, \lambda_k)du \le g(\delta_1, p_k, \lambda_k)\int_0^{\delta_1}\left[u - x - \frac{(2-2k)x}{k(\alpha_n+2)-2}\right]^2 du.$$

Clearly the integral on the right hand side is finite. From the definitions of p_k and λ_k,

$$g(\delta_1, p_k, \lambda_k) = \frac{(k\alpha_n x)^{k(\alpha_n+2)-1}}{\Gamma(k(\alpha_n+2)-1)}\delta_1^{-k(\alpha_n+2)}e^{-k\alpha_n x\delta_1^{-1}}.$$

By the Stirling approximation, as $\alpha_n \to \infty$,

$$\begin{aligned} &\frac{(k\alpha_n x)^{k(\alpha_n+2)-1}}{\Gamma(k(\alpha_n+2)-1)} \\ &= \left(\frac{k\alpha_n x}{k(\alpha_n+2)-2}\right)^{k(\alpha_n+2)-2} \frac{k\alpha_n x e^{k(\alpha_n+2)-2}}{\sqrt{2\pi[k(\alpha_n+2)-2]}}[1+o(1)] \\ &= O\left(x^{k\alpha_n} e^{k\alpha_n}\sqrt{\alpha_n}\right). \end{aligned} \tag{6.3}$$

Therefore,

$$\begin{aligned} g(\delta_1, p_k, \lambda_k) &= O\left(x^{k\alpha_n} e^{k\alpha_n} \delta_1^{-k\alpha_n} e^{-k\alpha_n x \delta_1^{-1}} \sqrt{\alpha_n}\right) \\ &= O\left(\left[\frac{x}{\delta_1}\exp\left(1-\frac{x}{\delta_1}\right)\right]^{k\alpha_n}\sqrt{\alpha_n}\right). \end{aligned}$$

This relation and $\delta_1 < x$ now readily implies that $g(\delta_1, p_k, \lambda_k) = o(1/\alpha_n)$, which in turn implies that

$$\int_0^{\delta_1} (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du = o\left(\frac{1}{\alpha_n}\right). \tag{6.4}$$

Now take $\delta_2 = x + \gamma/2$. Then,

$$\left|\int_{\delta_2}^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du\right| \le c\int_{\delta_2}^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)du.$$

But,

$$\begin{aligned} &\int_{\delta_2}^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)du \\ &= \frac{(k\alpha_n x)^{k(\alpha_n+2)-1}}{\Gamma(k(\alpha_n+2)-1)} \int_{\delta_2}^{\infty} (u-\mu_k)^2 \left(\frac{1}{u}\right)^{k(\alpha_n+2)} \exp\left(-\frac{k\alpha_n x}{u}\right) du. \end{aligned}$$

The integral on the right hand side is bounded above by

$$4\int_{\delta_2}^{\infty} \left(\frac{1}{u}\right)^{k(\alpha_n+2)-2} \exp\left(-\frac{k\alpha_n x}{u}\right) du.$$

By the change of variable, $v = k\alpha_n x/u$, we obtain

$$\begin{aligned} &\int_{\delta_2}^{\infty} \left(\frac{1}{u}\right)^{k(\alpha_n+2)-2} \exp\left(-\frac{k\alpha_n x}{u}\right) du \\ &= \left(\frac{1}{k\alpha_n x}\right)^{k(\alpha_n+2)-3} \int_0^{k\alpha_n x/\delta_2} v^{k(\alpha_n+2)-4}\exp(-v)dv. \end{aligned} \tag{6.5}$$

As a function of v, $v^{k(\alpha_n+2)-4}\exp(-v)$ is increasing in $v \le k(\alpha_n+2)-4$, and decreasing in $v \ge k(\alpha_n+2)-4$. Since $\delta_2 > x$, so $k\alpha_n x/\delta_2 < k(\alpha_n+2)-4$ for α_n sufficiently large. Therefore, for all $v \in [0, k\alpha_n x/\delta_2]$,

$$v^{k(\alpha_n+2)-4}\exp(-v) \le \left(\frac{k\alpha_n x}{\delta_2}\right)^{k(\alpha_n+2)-4}\exp\left(-\frac{k\alpha_n x}{\delta_2}\right).$$

Plugging the above inequality into (6.5), we obtain that

$$\begin{aligned}
&\int_{\delta_2}^{\infty}\left(\frac{1}{u}\right)^{k(\alpha_n+2)-2}\exp\left(-\frac{k\alpha_n x}{u}\right)du \\
&\le \left(\frac{1}{k\alpha_n x}\right)^{k(\alpha_n+2)-3}\left(\frac{k\alpha_n x}{\delta_2}\right)^{k(\alpha_n+2)-4}\frac{k\alpha_n x}{\delta_2}\exp\left(-\frac{k\alpha_n x}{\delta_2}\right) \\
&= \left(\frac{1}{\delta_2}\right)^{k(\alpha_n+2)-3}\exp\left(-\frac{k\alpha_n x}{\delta_2}\right).
\end{aligned}$$

From (6.3), we have

$$\begin{aligned}
&\int_{\delta_2}^{\infty}(u-\mu_k)^2 g(u; p_k, \lambda_k)du \\
&\le O\left(x^{k\alpha_n}e^{k\alpha_n}\sqrt{\alpha_n}\right)\cdot\left(\frac{1}{\delta_2}\right)^{k(\alpha_n+2)-3}\exp\left(-\frac{k\alpha_n x}{\delta_2}\right) \\
&= O\left(\left[\frac{x}{\delta_2}\exp\left(1-\frac{x}{\delta_2}\right)\right]^{k\alpha_n}\sqrt{\alpha_n}\right) = o\left(\frac{1}{\alpha_n}\right),
\end{aligned}$$

because $0 < x < \delta_2$ implies $0 < (x/\delta_2)\exp(1-x/\delta_2) < 1$. Hence,

$$\int_{\delta_2}^{\infty}(u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du = o\left(\frac{1}{\alpha_n}\right). \tag{6.6}$$

Finally, we shall show that

$$\int_{\delta_1}^{\delta_2}(u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du = o(1/\alpha_n).$$

By uniform continuity of l'',

$$\begin{aligned}
&\int_{\delta_1}^{\delta_2}(u-\mu_k)^2 g(u; p_k, \lambda_k)|l''(\tilde{u}) - l''(\mu_k)|du \\
&\le \epsilon\int_0^{\infty}(u-\mu_k)^2 g(u; p_k, \lambda_k)du,
\end{aligned}$$

by the fact that $|\tilde{u}-\mu_k| \le |u-\mu_k| < \gamma$, for $u \in [\delta_1, \delta_2]$ and α_n sufficiently large. Because

$\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)du = O(1/\alpha_n)$, we obtain

$$\int_{\delta_1}^{\delta_2} (u-\mu_k)^2 g(u;p_k,\lambda_k)|l''(\tilde{u}) - l''(\mu_k)|du = \epsilon \cdot O\left(\frac{1}{\alpha_n}\right). \tag{6.7}$$

The arbitrariness of ϵ combined with (6.4), (6.6) and (6.7), finally yield

$$\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du = o\left(\frac{1}{\alpha_n}\right).$$

Hence the desired result in the Lemma. □

In particular, if $k=1$, then

$$\int_0^\infty g(u;p_1,\lambda_1)l(u)du = l(x) + \frac{x^2 l''(x)}{2(\alpha_n - 1)} + o\Big(\frac{1}{\alpha_n}\Big). \tag{6.8}$$

To analyze the limiting behavior of $\hat{f}_n(x)$ as $x \to 0$, similar to the symmetric kernel case, we analyze the limiting bias of $\hat{f}_n(x)$ at $x = u/\alpha_n$, where $0 < u < 1$. It is easy to see that

$$\hat{f}_n\left(\frac{u}{\alpha_n}\right) = \frac{1}{n}\sum_{i=1}^n \frac{1}{X_i \Gamma(\alpha_n+1)}\left(\frac{u}{X_i}\right)^{\alpha_n+1} e^{-u/X_i}.$$

Let $p = \alpha_n + 1$, $\lambda = u$, we can show that

$$E\hat{f}_n\left(\frac{u}{\alpha_n}\right) = \int_0^\infty g(x,p,\lambda)f(x)dx = f\left(\frac{u}{\alpha_n}\right) + O\left(\frac{1}{\alpha_n}\right).$$

Therefore, $\hat{f}_n(x)$ does not suffer from the boundary effect.

The following decomposition of $\hat{m}_n(x)$ will be used repeatedly in the proofs below.

$$\hat{m}_n(x) - m(x) = \frac{B_n(x)+V_n(x)}{f(x)} + \Big[\frac{1}{\hat{f}_n(x)} - \frac{1}{f(x)}\Big][B_n(x)+V_n(x)],$$

where

$$B_n(x) = \frac{1}{n}\sum_{i=1}^n K_{\alpha_n}(x,X_i)[m(X_i)-m(x)], \quad V_n(x) = \frac{1}{n}\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_i,$$

with $K_{\alpha_n}(x,X_i)$ defined in (2.4). Now we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. First, we shall compute the conditional bias of $\hat{m}_n(x)$. Direct calculations shows that $E[\hat{m}_n(x)|\mathbf{X}] - m(x) = B_n(x)/\hat{f}_n(x)$. Since $\hat{f}_n(x) = f(x) + o_p(1)$, it suffices to discuss the asymptotic property of $B_n(x)$. Note that $EB_n(x) =$

$EK_{\alpha_n}(x,X)m(X) - m(x)EK_{\alpha_n}(x)$. But

$$\begin{aligned} E(K_{\alpha_n}(x,X)m(X)) &= \int_0^\infty \frac{1}{u}\Big(\frac{\alpha_n x}{u}\Big)^{\alpha_n+1}\frac{\exp(-\alpha_n x/u)}{\Gamma(\alpha_n+1)}m(u)f(u)du \\ &= \int_0^\infty g(u,p_1,\lambda_1)m(u)f(u)du, \end{aligned}$$

where $p_1 = \alpha_n + 1$, $\lambda_1 = \alpha_n x$. Let $H(u) = m(u)f(u)$. Applying (6.8) with $l(u) = H(u)$ and with $l(u) = f(u)$, respectively, yields

$$\begin{aligned} E(K_{\alpha_n}(x,X)m(X_1)) &= m(x)f(x) + \frac{x^2H''(x)}{2(\alpha_n-1)} + o\Big(\frac{1}{\alpha_n}\Big), \\ m(x)EK_{\alpha_n}(x,X) &= m(x)\left[f(x) + \frac{x^2f''(x)}{2(\alpha_n-1)}\right] + o\Big(\frac{1}{\alpha_n}\Big). \end{aligned}$$

Therefore,

$$EB_n(x) = \frac{x^2[H''(x) - m(x)f''(x)]}{2(\alpha_n-1)} + o\Big(\frac{1}{\alpha_n}\Big). \qquad (6.9)$$

Direct calculations show that $x^2[H''(x)-m(x)f''(x)]/2 = b(x)f(x)$, where $b(x)$ is defined in (3.1).

Next, consider

$$\mathrm{Var}(B_n(x)) = \frac{1}{n}EK^2_{\alpha_n}(x,X)[m(X)-m(x)]^2 - \frac{1}{n}[EK_{\alpha_n}(x,X)(m(X)-m(x))]^2.$$

Note that $EK^2_{\alpha_n}(x,X)(m(X)-m(x))^2$ equals

$$\begin{aligned} &\int_0^\infty \frac{1}{u^2}\Big(\frac{\alpha_n x}{u}\Big)^{2(\alpha_n+1)}\frac{1}{\Gamma(\alpha_n+1)}\exp\Big(-\frac{2\alpha_n x}{u}\Big)(m(u)-m(x))^2f(u)du \\ &= \frac{\Gamma(2\alpha_n+3)}{x\alpha_n 2^{2\alpha_n+3}\Gamma^2(\alpha_n+1)}\int_0^\infty g(u;p_2,\lambda_2)(m(u)-m(x))^2f(u)du, \end{aligned}$$

where $p_2 = 2\alpha_n + 3$, $\lambda_2 = 2\alpha_n x$. By the Stirling approximation, for α_n sufficiently large,

$$\frac{\Gamma(2\alpha_n+3)}{\alpha_n 2^{2\alpha_n+3}\Gamma^2(\alpha_n+1)} = \frac{\sqrt{\alpha_n}}{2\sqrt{\pi}}[1+o(1)].$$

A Taylor expansion of $m(u)$ and $f(u)$ around $\alpha_n x/(\alpha_n+1)$ up to the first order gives the following expansion for $\int_0^\infty g(u;p_2,\lambda_2)(m(u)-m(x))^2f(u)du$.

$$(m'(x))^2f(x)\int_0^\infty \Big(u - \frac{\alpha_n x}{\alpha_n+1}\Big)^2 g(u;p_2,\lambda_2)du + o\Big(\frac{1}{\alpha_n}\Big),$$

by the assumptions (A1) and (A2), and the fact

$$\int_0^\infty \left(u - \frac{\alpha_n x}{\alpha_n + 1}\right)^2 g(u; p_2, \lambda_2) du = \frac{x^2 \alpha_n^2}{(\alpha_n + 1)^2 (2\alpha_n + 1)} = O\left(\frac{1}{\alpha_n}\right).$$

Therefore,

$$\frac{1}{n} E K_{\alpha_n}^2 (x, X)[m(X) - m(x)]^2 = O\left(\frac{1}{n\sqrt{\alpha_n}}\right)$$

From (6.9), $EB_n(x) = O(1/\alpha_n)$. Hence

$$\mathrm{Var}(B_n(x)) = O\left(\frac{1}{n\sqrt{\alpha_n}}\right) + O\left(\frac{1}{n\alpha_n^2}\right). \tag{6.10}$$

Therefore, (6.9), (6.10), and the fact $x^2[H''(x) - m(x)f''(x)]/2 = b(x)f(x)$ together yield

$$\frac{B_n(x)}{f(x)} = \frac{b(x)}{\alpha_n - 1} + o_p\left(\frac{1}{\alpha_n}\right) + O_p\left(\frac{1}{\sqrt{n\sqrt{\alpha_n}}}\right). \tag{6.11}$$

Moreover,

$$\begin{aligned} & E[\hat{m}_n(x)|\mathbf{X}] - m(x) \\ & = [1/f(x) + o_p(1)] \cdot [EB_n(x) + B_n(x) - EB_n(x)] \\ & = \left[\frac{1}{f(x)} + o_p(1)\right] \cdot \left[\frac{b(x)f(x)}{\alpha_n} + o_p\left(\frac{1}{\alpha_n}\right) + O_p\left(\frac{1}{n\sqrt{\alpha_n}}\right)\right], \end{aligned}$$

which implies the claim (3.2) about the conditional bias of $\hat{m}_n(x)$.

Next, we verify the claim (3.3) about the conditional variance of $\hat{m}_n(x)$. In fact, with $\sigma^2(x) = E(\varepsilon^2 | X = x)$,

$$\mathrm{Var}[\hat{m}_n(x)|\mathbf{X}] = \frac{1}{\hat{f}_n^2(x)} \cdot \frac{1}{n^2} \sum_{i=1}^n K_{\alpha_n}^2 \left(\frac{x}{X_i}\right) \sigma^2(X_i). \tag{6.12}$$

Verify that under condition (A3) about $\sigma^2(x)$,

$$E\left[\frac{1}{n^2} \sum_{i=1}^n K_{\alpha_n}^2 (x, X_i) \sigma^2(X_i)\right] = \frac{\sigma^2(x) f(x) \sqrt{\alpha_n}}{2nx\sqrt{\pi}} + o\left(\frac{\sqrt{\alpha_n}}{n}\right),$$

which, together with (6.12) and the fact $\hat{f}_n(x) = f(x) + o_p(1)$, implies the claim (3.3).

□

Proof of Theorem 3.2. Let $\xi_{in}(x) = n^{-1}[K_{\alpha_n}(x, X_i) - EK_{\alpha_n}(x, X)]$. Then

$$\hat{f}_n(x) = \sum_{i=1}^{n} \xi_{in}(x) + EK_{\alpha_n}(x, X).$$

Since $EK_{\alpha_n}(x, X) = f(x) + x^2 f''(x)/2(\alpha_n - 1) + o(1/\alpha_n)$,

$$\hat{f}_n(x) - f(x) - \frac{x^2 f''(x)}{2(\alpha_n - 1)} + o\Big(\frac{1}{\alpha_n}\Big) = \sum_{i=1}^{n} \xi_{in}(x).$$

Lindeberg-Feller CLT will be used to show the asymptotic normality of $\sum_{i=1}^{n} \xi_{in}(x)$. For any $a > 0, b > 0$ and $r > 1$, using the well known inequality $(a+b)^r \le 2^{r-1}(a^r + b^r)$, we have

$$E|\xi_{in}(x)|^{2+\delta} \le n^{-(2+\delta)} 2^{1+\delta} [E(K_{\alpha_n}(x, X))^{2+\delta} + (EK_{\alpha_n}(x, X))^{2+\delta}].$$

Let $\lambda_\delta = (2+\delta)\alpha_n x$, $p_\delta = (2+\delta)(2+\alpha_n) - 1$. A tedious calculation shows that $E(K_{\alpha_n}(x, X))^{2+\delta}$ can be written as

$$\frac{1}{(\alpha_n x)^{1+\delta}(2+\delta)^{(2+\delta)(2+\alpha_n)-1}} \frac{\Gamma((2+\delta)(2+\alpha_n)-1)}{\Gamma^{2+\delta}(\alpha_n+1)} \int_0^\infty g(u; p_\delta, \lambda_\delta) f(u) du.$$

For n and α_n, large enough, using the Stirling approximation, we have

$$\frac{\Gamma((2+\delta)(2+\alpha_n)-1)}{\Gamma^{2+\delta}(\alpha_n+1)} = O\Big((2+\delta)^{(2+\delta)(2+\alpha_n)} \alpha_n^{2(2+\delta)-(5+\delta)/2}\Big).$$

Also, we have $\int_0^\infty g(u; p_\delta, \lambda_\delta) f(u) du = f(x) + o(1)$. Hence

$$EK_{\alpha_n}^{2+\delta}(x, X) = O\Big(\alpha_n^{(\delta+1)/2}\Big).$$

Note that

$$EK_{\alpha_n}(x, X) = \int_0^\infty g(u; p_1, \lambda_1) f(u) du,$$
$$EK_{\alpha_n}^2(x, X) = \frac{\Gamma(2\alpha_n + 3)}{x\alpha_n 2^{2\alpha_n+3}\Gamma^2(\alpha_n+1)} \int_0^\infty g(u; p_2, \lambda_2) f(u) du.$$

Hence, by Lemma 6.1, we obtain

$$
\begin{aligned}
v_n^2 &= \mathrm{Var}\Big(\sum_{i=1}^{n} \xi_{in}(x)\Big) = \mathrm{Var}(\hat{f}_n(x)) \qquad (6.13)\\
&= n^{-1}\big[EK_{\alpha_n}^2(x, X) - (EK_{\alpha_n}(x, X))^2\big]\\
&= \frac{\sqrt{\alpha_n} f(x)}{2nx\sqrt{\pi}} + o\Big(\frac{\sqrt{\alpha_n}}{n}\Big).
\end{aligned}
$$

This fact together with $EK_{\alpha_n}(x, X) = f(x) + o(1)$, imply

$$
v_n^{-(2+\delta)} \sum_{i=1}^{n} E\xi_{in}^{2+\delta}(x) = nv_n^{-(2+\delta)} E\xi_{1n}^{2+\delta} = O\Big(\Big(\frac{\sqrt{\alpha_n}}{n}\Big)^{\delta/2}\Big),
$$

which converges to 0, by assumption (A4). Hence the Lindeberg-Feller condition holds. This completes the proof of the Theorem 3.2. □

PROOF OF THEOREM 3.3. Fix an $x > 0$. To show the asymptotic normality of $\hat{m}_n(x)$, again we use the decomposition (6.13).

We shall first show that $V_n(x)$ is asymptotically normal. For this purpose, let $\eta_{in} = n^{-1}K_{\alpha_n}(x, X_i)\varepsilon_i$ so that $V_n(x) = \sum_{i=1}^{n} \eta_{in}$. Clearly, $E\eta_{in} = 0$. By assumption (A3) on $\sigma^2(x)$, a routine argument leads to $E\eta_{in}^2 = [\sqrt{\alpha_n} f(x)\sigma^2(x)/(2n^2x\sqrt{\pi})][1+o(1)]$. Therefore,

$$
s_n^2 = \mathrm{Var}\Big(\sum_{i=1}^{n} \eta_{in}\Big) = nE\eta_{in}^2 = \frac{f(x)\sigma^2(x)\sqrt{\alpha_n}}{2nx\sqrt{\pi}}[1 + o(1)].
$$

Using a similar argument as in dealing with $E|\xi_{in}(x)|^{2+\delta}$ in the proof of Theorem 3.2, verify that for any $\delta > 0$,

$$
E|\eta_{in}|^{2+\delta} = n^{-(2+\delta)} EK_{\alpha_n}^{2+\delta}(x, X)E(|\varepsilon|^{2+\delta}|X = x) = O(n^{-(2+\delta)}\alpha_n^{(1+\delta)/2}).
$$

Hence

$$
s_n^{-(2+\delta)} \sum_{i=1}^{n} E|\eta_{in}|^{2+\delta} = O\Big(\Big(\frac{\sqrt{\alpha_n}}{n}\Big)^{\delta/2}\Big) = o(1).
$$

Hence, by the Lindeberg-Feller CLT, $s_n^{-1}V_n(x) \to_d N(0, 1)$.

From the asymptotic results on $\hat{f}_n(x)$ and $V_n(x)$ in Theorem 3.2 and fact (6.11) about $B_n(x)$, we obtain that

$$
s_n^{-1}\Big[\frac{1}{\hat{f}_n(x)} - \frac{1}{f(x)}\Big][B_n(x) + V_n(x)] = o_p(1).
$$

This, together with the result that $\sqrt{n/\sqrt{\alpha_n}} \cdot O_p(1/\sqrt{n\sqrt{\alpha_n}}) = o_p(1)$, implies

$$f(x)s_n^{-1}\Big(\hat{m}_n(x) - m(x) - \frac{b(x)}{\alpha_n - 1} + o\Big(\frac{1}{\alpha_n}\Big)\Big) = s_n^{-1}V_n(x) \to_d N(0,1).$$

The proof is completed by noticing that $f(x)s_n^{-1} = \Big(v(x)\sqrt{\alpha_n}/n\Big)^{-1/2}$. □

Proof of Theorem 3.4. Recall that $E\hat{f}_n(x) = \int_0^\infty g(u; p_1, \lambda_1)f(u)du$. By (6.8) and the boundedness of $x^2 f''(x)$ on $[a,b]$, we obtain

$$E\hat{f}_n(x) - f(x) = O\Big(\frac{1}{\alpha_n}\Big), \quad \text{for any } x \in [a,b].$$

Hence $\sup_{a\le x\le b}|E\hat{f}_n(x) - f(x)| = O(1/\alpha_n)$. So, we only need to show that $\hat{f}_n(x) - E\hat{f}_n(x) = o(\alpha_n^{1/4}\sqrt{\log n}/\sqrt{n})$. For this purpose, let $\xi_{in}(x) = n^{-1}[K_{\alpha_n}(x, X_i) - EK_{\alpha_n}(x, X_i)]$, hence $\hat{f}_n(x) - E\hat{f}_n(x) = \sum_{i=1}^n \xi_{in}(x)$. In order to apply Bernstein inequality, we have to verify the Cramér condition for ξ_{in}, that is, we need to show that, for $k \ge 3$, $E|\xi_{1n}|^k \le c_n^{k-2}k!E\xi_{1n}^2$ for some c_n only depending on n.

Note that $K_{\alpha_n}(x, X)$ can be written as

$$K_{\alpha_n}(x, X) = \frac{\alpha_n^{\alpha_n+1}}{x\Gamma(\alpha_n+1)}\left(\frac{x}{X}\right)^{\alpha_n+2}\exp\left(-\frac{\alpha_n x}{X}\right).$$

As a function of u, $u^{\alpha_n+2}\exp(-\alpha_n u)$ attains its maximum at $u = (\alpha_n+2)/\alpha_n$. Therefore, for any x and X, by Stirling formula,

$$\begin{aligned}
K_{\alpha_n}(x, X) &\le \frac{\alpha_n^{\alpha_n+1}}{x\Gamma(\alpha_n+1)}\left(\frac{\alpha_n+2}{\alpha_n}\right)^{\alpha_n+2}\exp\left(-(\alpha_n+2)\right)\\
&\le \frac{(\alpha_n+2)^2}{x\alpha_n}\frac{(\alpha_n+2)^{\alpha_n}}{\Gamma(\alpha_n+1)}\exp\left(-(\alpha_n+2)\right)\\
&= \frac{(\alpha_n+2)^2}{x\alpha_n}\frac{(\alpha_n+2)^{\alpha_n}}{\alpha_n^{\alpha_n}\sqrt{2\pi\alpha_n}e^{-\alpha_n}(1+o(1))}\exp\left(-(\alpha_n+2)\right)\\
&\le \frac{c\sqrt{\alpha_n}}{x},
\end{aligned} \tag{6.14}$$

for some positive constant c. Therefore, for any $k \geq 3$, and α_n large enough,

$$\begin{aligned} E|\xi_{in}|^k &= n^{-k} E|K_{\alpha_n}(x, X_i) - EK_{\alpha_n}(x, X_i)|^k \\ &\leq \left(\frac{c\sqrt{\alpha_n}}{xn}\right)^{k-2} n^{-2} E|K_{\alpha_n}(x, X_i) - EK_{\alpha_n}(x, X_i)|^2 \\ &= \left(\frac{c\sqrt{\alpha_n}}{xn}\right)^{k-2} E\xi_{in}^2. \end{aligned}$$

With $v_n := \left(\sum_{i=1}^n E\xi_{in}^2\right)^{1/2}$, this immediately implies,

$$E|\xi_{in}|^k \leq k!\Big(\frac{c\sqrt{\alpha_n}}{nx}\Big)^{k-2} E\xi_{in}^2, \quad \forall\ 1 \leq i \leq n,$$

or

$$E\Big(\frac{\xi_{in}}{v_n}\Big)^k \leq k!\Big(\frac{c\sqrt{\alpha_n}}{nxv_n}\Big)^{k-2} E\Big[\frac{\xi_{in}}{v_n}\Big]^2 \quad \forall\ 1 \leq i \leq n.$$

By (6.13), $v_n^2 = \sqrt{\alpha_n} f(x)/2nx\sqrt{\pi} + o\Big(\sqrt{\alpha_n}/n\Big)$. This together with the fact that $xf(x)$ is bounded away from 0 and ∞ on $[a, b]$, implies

$$E\Big[\frac{\xi_{in}}{v_n}\Big]^k \leq k!\Big(\frac{c{\alpha_n}^{1/4}}{\sqrt{n}}\Big)^{k-2} E\Big[\frac{\xi_{in}}{v_n}\Big]^2. \tag{6.15}$$

Then, by (6.15) and the Bernstein inequality, for any positive number c,

$$P\Bigg(\Big|\frac{\sum_{i=1}^n \xi_{in}}{v_n}\Big| \geq c\sqrt{\log n}\Bigg) \leq 2\exp\Bigg(-\frac{c^2 \log n}{4\Big(1 + c\alpha_n^{1/4}\sqrt{\log n}/\sqrt{n}\Big)}\Bigg).$$

Since $\alpha_n^{1/2} \log n/n \to 0$, so for n large enough,

$$P\Bigg(\Big|\frac{\sum_{i=1}^n \xi_{in}}{v_n}\Big| \geq c\sqrt{\log n}\Bigg) \leq 2\exp\Big(-\frac{c^2 \log n}{8}\Big).$$

Upon taking $c = 8$, we have

$$P\Big(\Big|\sum_{i=1}^n \xi_{in}\Big| \geq c\sqrt{\log n} v_n\Big) = \frac{2}{n^8}.$$

Since $\sum_{n=1}^\infty n^{-8} < \infty$, so by the Borel-Cantelli Lemma and by the fact $v_n^2 = O(\sqrt{\alpha_n}/n)$, we obtain

$$\hat{f}_n(x) - E\hat{f}_n(x) = \sum_{i=1}^{n} \xi_{in} = o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big).$$

To bound the $\sum_{i=1}^{n} \xi_{in}$ uniformly for all $x \in [a,b]$, we partition the interval $[a,b]$ by the equally spaced points x_i, $i = 0,1,2,\ldots,N_n$, such that $a = x_0 < x_1 < x_2 < \cdots < x_{N_n} = b$, $N_n = n^3$. It is easily seen that

$$P\Big(\max_{0\le j\le N_n}\Big|\sum_{i=1}^{n}\xi_{in}(x_j)\Big| > c\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big) \le \frac{2N_n}{n^8} = \frac{2}{n^5}.$$

Borel-Cantelli Lemma implies that

$$\max_{0\le j\le N_n}\Big|\sum_{i=1}^{n}\xi_{in}(x_j)\Big| = o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big). \tag{6.16}$$

For any $x \in [x_j, x_{j+1}]$,

$$\begin{aligned}\xi_{in}(x) - \xi_{in}(x_j) = n^{-1}&[K_{\alpha_n}(x,X_i) - EK_{\alpha_n}(x,X_i)]\\ &-n^{-1}[K_{\alpha_n}(x_j,X_i) - EK_{\alpha_n}(x_j,X_i)].\end{aligned}$$

Then a Taylor expansion of $K_{\alpha_n}(x,X_i)$ at $x = x_j$ up to the first order leads to the following expression for the difference $K_{\alpha_n}(x,X_i) - K_{\alpha_n}(x_j,X_i)$:

$$\begin{aligned}\frac{x-x_j}{\Gamma(\alpha_n+1)\alpha_n\tilde{x}^2}\Big[(\alpha_n+1)\Big(\frac{\alpha_n\tilde{x}}{X_i}\Big)^{\alpha_n+2}&\exp\Big(-\frac{\alpha_n\tilde{x}}{X_i}\Big)\\ -\Big(\frac{\alpha_n\tilde{x}}{X_i}\Big)^{\alpha_n+3}&\exp\Big(-\frac{\alpha_n\tilde{x}}{X_i}\Big)\Big],\end{aligned}$$

where $|x-\tilde{x}| \le x_{j+1} - x_j \le (b-a)/N_n$. Note that for $p > 0$, the maximum of $x^p e^{-x}$ for $x > 0$ is attained at $x = p$ and equals $p^p e^{-p}$. Hence,

$$\Big(\frac{\alpha_n\tilde{x}}{X_i}\Big)^{\alpha_n+2}\exp\Big(-\frac{\alpha_n\tilde{x}}{X_i}\Big) \le (\alpha_n+2)^{\alpha_n+2}e^{-\alpha_n-2},$$

$$\Big(\frac{\alpha_n\tilde{x}}{X_i}\Big)^{\alpha_n+3}\exp\Big(-\frac{\alpha_n\tilde{x}}{X_i}\Big) \le (\alpha_n+3)^{\alpha_n+3}e^{-\alpha_n-3}.$$

Therefore, for al $1 \le i \le n$,

$$
\begin{aligned}
&|K_{\alpha_n}(x, X_i) - K_{\alpha_n}(x_j, X_i)| \\
&\le \frac{(x - x_j)\alpha_n^{\alpha_n+2}\exp(-\alpha_n)}{\Gamma(\alpha_n+1)\tilde{x}^2}\Big[\Big(1 + \frac{2}{\alpha_n}\Big)^{\alpha_n+3} e^{-2} + \Big(1 + \frac{3}{\alpha_n}\Big)^{\alpha_n+3} e^{-3}\Big].
\end{aligned}
$$

This upper bound together with the Stirling approximation for the Gamma function, one concludes that for n and α_n large enough,

$$|K_{\alpha_n}(x, X_i) - K_{\alpha_n}(x_j, X_i)| \le c(x - x_j)\alpha_n^{3/2}/\tilde{x}^2,$$

for some positive constant c. Because $0 \le x - x_j \le (b - a)/N_n$, and $\tilde{x} > 1/a$,

$$|K_{\alpha_n}(x, X_i) - K_{\alpha_n}(x_j, X_i)| \le \frac{c\alpha_n^{3/2}}{N_n},$$

which implies that when n is large enough, for some constant c,

$$|\xi_{in}(x) - \xi_{in}(x_j)| \le \frac{c\alpha_n^{3/2}}{nN_n}, \quad 1 \le i \le n. \tag{6.17}$$

These bounds imply that for all $x \in [x_j, x_{j+1}]$ and $0 \le j \le N_n - 1$,

$$\Big|\sum_{i=1}^{n} \xi_{in}(x) - \sum_{i=1}^{n} \xi_{in}(x_j)\Big| \le \frac{c\alpha_n^{3/2}}{n^3} = o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big). \tag{6.18}$$

Finally, from (6.16) and (6.18), we obtain

$$
\begin{aligned}
&\sup_{a \le x \le b} |\hat{f}_n(x) - E\hat{f}_n(x)| = \sup_{a \le x \le b} \Big|\sum_{i=1}^{n} \xi_{in}(x)\Big| \\
&\le \max_{0 \le j \le N_n} \Big|\sum_{i=1}^{n} \xi_{in}(x_j)\Big| + \max_{0 \le j \le N_n - 1} \sup_{x \in [x_j, x_{j+1}]} \Big|\sum_{i=1}^{n} \xi_{in}(x) - \sum_{i=1}^{n} \xi_{in}(x_j)\Big| \\
&= o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big).
\end{aligned}
$$

This, together with the result $\sup_{a \le x \le b} |E\hat{f}_n(x) - f(x)| = O(1/\alpha_n)$, completes the proof of Theorem 3.4. □

Proof of Theorem 3.5. By (6.13) and Theorem 3.4, it suffices to prove the following two facts:

$$\sup_{x\in[a,b]}\Big|\frac{B_n(x)}{\hat{f}_n(x)}\Big| = O\Big(\frac{1}{\alpha_n}\Big) + o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big), \tag{6.19}$$

$$\sup_{x\in[a,b]}\Big|\frac{V_n(x)}{\hat{f}_n(x)}\Big| = O\Big(\frac{1}{\alpha_n}\Big) + o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big). \tag{6.20}$$

We shall prove (6.20) only, the proof of (6.19) being similar.

Let β, η be such that $\beta < 2/5$, $\beta(2+\eta) > 1$ and $\beta(1+\eta) > 2/5$ and define $d_n = n^\beta$. For each i, write $\varepsilon_i = \varepsilon_{i1}^{d_n} + \varepsilon_{i2}^{d_n} + \mu_i^{d_n}$, with

$$\varepsilon_{i1}^{d_n} = \varepsilon_i I(|\varepsilon_i| > d_n),\ \varepsilon_{i2}^{d_n} = \varepsilon_i I(|\varepsilon_i| \le d_n) - \mu_i^{d_n},\ \mu_i^{d_n} = E[\varepsilon_i I(|\varepsilon_i| \le d_n)|X_i].$$

Hence,

$$\frac{V_n(x)}{\hat{f}_n(x)} = \frac{\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_{i1}^{d_n}}{\sum_{i=1}^n K_{\alpha_n}(x,X_i)} + \frac{\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_{i2}^{d_n}}{\sum_{i=1}^n K_{\alpha_n}(x,X_i)} + \frac{\sum_{i=1}^n K_{\alpha_n}(x,X_i)\mu_i^{d_n}}{\sum_{i=1}^n K_{\alpha_n}(x,X_i)}.$$

Since $E(\varepsilon_i|X_i) = 0$, so $\mu_i^{d_n} = -E[\varepsilon_i I(|\varepsilon_i| > d_n)|X_i]$, then from assumption (A4), we have $|\mu_i^{d_n}| \le c d_n^{-(1+\eta)}$. Hence

$$\sup_{x\in[a,b]}\Big|\frac{\sum_{i=1}^n K_{\alpha_n}(x,X_i)\mu_i^{d_n}}{\sum_{i=1}^n K_{\alpha_n}(x,X_i)}\Big| \le c d_n^{-(1+\eta)} = o\Big(\frac{\alpha_n^{1/4}}{\sqrt{n}}\Big).$$

Now, consider the part involving $\varepsilon_{i1}^{d_n}$. By the Markov inequality,

$$\sum_{n=1}^\infty P(|\varepsilon_n| > d_n) \le E|\varepsilon|^{2+\eta}\sum_{n=1}^n \frac{1}{d_n^{2+\eta}} < \infty.$$

Borel-Cantelli Lemma implies that

$$\begin{aligned} &P\{\exists N, |\varepsilon_n| \le d_n \text{ for } n > N\} = 1 \\ \Rightarrow\ &P\{\exists N, |\varepsilon_i| \le d_n, i = 1,2,\ldots,n, \text{ for } n > N\} = 1 \\ \Rightarrow\ &P\{\exists N, \varepsilon_{i,1}^{d_n} = 0, i = 1,2,\ldots,n, \text{ for } n > N\} = 1. \end{aligned}$$

Hence,

$$\sup_{x\in[a,b]}\Big|\frac{\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_{i,1}^{d_n}}{\sum_{i=1}^n K_{\alpha_n}(x,X_i)}\Big| = O(n^{-k}), \quad \forall k > 0.$$

For the term $\varepsilon_{i,2}^{d_n}$, we have $E[\varepsilon_{i,2}^{d_n}|X_i] = 0$, and it is easy to show that

$$\text{Var}(\varepsilon_{i,2}^{d_n}|X_i) = \sigma^2(X_i) + O[d_n^{-\eta} + d_n^{-2(1+\eta)}]$$

and for $k \geq 2$, $E(|\varepsilon_{i,n}^{d_n}|^k|X_i) \leq 2^{k-2}d_n^{k-2}E(|\varepsilon_{i,n}^{d_n}|^2|X_i)$. Then from (6.14) and the boundedness of $\sigma^2(x)$ over $(0,\infty)$, we have

$$\begin{aligned}
E|n^{-1}K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}|^k &\leq n^{-k}E[K_{\alpha_n}^k(x,X)E(|\varepsilon_{i,n}^{d_n}|^k|X_i)] \\
&\leq cn^{-k}2^{k-2}d_n^{k-2}EK_{\alpha_n}^k(x,X)\sigma^2(X) \\
&\leq \Big(cd_n\sqrt{\alpha_n}/n\Big)^{k-2}E|n^{-1}K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}|^2.
\end{aligned}$$

Because

$$\begin{aligned}
E|n^{-1}K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}|^2 &= \frac{1}{n^2}E[K_{\alpha_n}^2(x,X)\sigma^2(X)][1+o(1)] \\
&= \frac{\sqrt{\alpha_n}f(x)\sigma^2(x)}{2n^2\sqrt{\pi}x}[1+o(1)],
\end{aligned}$$

the r.v. $n^{-1}K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}$ satisfies the Cramér condition. So, using the Bernstein inequality as in proving Theorem 3.4, one establishes the fact that for all $c > 0$,

$$\begin{aligned}
P\Big(\Big|\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}\Big| \geq c\sqrt{\log n}\sqrt{\sum_{i=1}^n E\Big[K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}\Big]^2}\Big) \\
\leq 2\exp(-c^2\log n/8).
\end{aligned}$$

Take $c = 4$ and $C(x) = c\sqrt{f(x)\sigma^2(x)/(2x\sqrt{\pi})}$ in the above inequality to obtain

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}\Big| \geq C(x)\sqrt{\alpha_n^{1/2}\log n/n}\Big) \leq \frac{2}{n^2},$$

by Borel-Cantelli Lemma and the boundedness of $f(x)\sigma^2(x)/x$ over $x \in [a,b]$, this implies, for each $x \in [a,b]$,

$$\Big|\frac{1}{n}\sum_{i=1}^n K_{\alpha_n}(x,X_i)\varepsilon_{i,2}^{d_n}\Big| = o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big).$$

To show the above bound is indeed uniform, we can use the similar technique as in showing the uniform convergence of $\hat{f}_n(x)$ as in the proof of Theorem 3.4. In fact, the only major difference is that, instead of using (6.17), we should use the inequality

$$\Big|K_{\alpha_n}(x, X_i)\varepsilon_{i,2}^{d_n} - K_{\alpha_n}(x_j, X_i)\varepsilon_{i,2}^{d_n}\Big| \le \frac{c\alpha_n^{3/2} d_n}{N_n}, \quad x \in [x_j, x_{j+1}],\, 1 \le i \le n.$$

The above result, together with the facts that $f(x)$ is bounded below from 0 on $[a, b]$, and $\sup_{x\in[a,b]} |\hat{f}_n(x) - f(x)| = o(1)$, implies

$$\sup_{x\in[a,b]} \Big|\frac{\sum_{i=1}^n K_{\alpha_n}(x, X_i)\varepsilon_{i,2}^{d_n}}{\sum_{i=1}^n K_{\alpha_n}(x, X_i)}\Big| = o\Big(\frac{\alpha_n^{1/4}\sqrt{\log n}}{\sqrt{n}}\Big), \quad \text{a.s.}$$

This concludes the proof of Theorem 3.5.

Acknowledgement

The authors gratefully acknowledge the editors and two referees for their helpful comments which improved the presentation of the paper. The research is supported by the NSF DMS grant 1205271, 1205276.

References

Abadir, K.M., and Lawford, S. (2004), “Optimal asymmetric kernels,” *Economics Letters*, 83, 61–68.

Bouezmarni, T., and Rolin, J. (2003), “Consistency of the beta kernel density function estimator,” *Canadian Journal of Statistics*, 31, 89–98.

Chaubey, Y.P., Sen, A., and Sen, P.K. (2012), “A new smooth density estimator for non-negative random variables,” *Journal of India Statistical Association*, 50, 22.

Chen, S.X. (1999), “Beta kernel estimators for density functions,” *Computational Statistics & Data Analysis*, 31, 131–145.

Chen, S.X. (2000a), “Beta kernel smoothers for regression curves,” *Statistica Sinica*, 10, 73–91.

Chen, S.X. (2000b), “Probability density function estimation using gamma kernels,” *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

Chen, S.X. (2002), “Local linear smoothers using asymmetric kernels,” *Annals of the Institute of Statistical Mathematics*, 54, 312–323.

Cline, D.B. (1988), "Admissibile kernel estimators of a multivariate density," *The Annals of Statistics*, 16, 1421–1427.

Cowling, A., and Hall, P. (1996), "On pseudodata methods for removing boundary effects in kernel density estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 551–563.

Fan, J. (1993), "Local linear regression smoothers and their minimax efficiencies," *The Annals of Statistics*, pp. 196–216.

Fan, J., and Gijbels, I. (1992), "Variable bandwidth and local linear regression smoothers," *The Annals of Statistics*, pp. 2008–2036.

Gasser, T., and Müller, H.G. (1979), "Kernel estimation of regression functions," *Smoothing techniques for curve estimation*, pp. 23–68.

Härdle, W., Hall, P., and Marron, J. (1992), "Regression smoothing parameters that are not far from their optimum," *Journal of the American Statistical Association*, 87, 227–233.

Härdle, W., Hall, P., and Marron, J.S. (1988), "How far are automatically chosen regression smoothing parameters from their optimum?," *Journal of the American Statistical Association*, 83, 86–95.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A., *Nonparametric and semiparametric models*, Springer Verlag (2004).

Hart, J.D., *Nonparametric smoothing and lack-of-fit tests*, Springer New York (1997).

John, R. (1984), "Boundary modification for kernel regression," *Communications in Statistics-Theory and Methods*, 13, 893–900.

Jones, M. (1993), "Simple boundary correction for kernel density estimation," *Statistics and Computing*, 3, 135–146.

Jones, M., and Henderson, D. (2007), "Miscellanea kernel-type density estimation on the unit interval," *Biometrika*, 94, 977–984.

Kotz, S., Balakrishnan, N., and Johnson, N.L., *Continuous multivariate distributions, models and applications*, Vol. 334, Wiley-Interscience (2000).

Marron, J.S., and Ruppert, D. (1994), "Transformations to reduce boundary bias in kernel density estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 653–671.

Mnatsakanov, R., and Sarkisian, K. (2012), "Varying kernel density estimation on," *Statistics & Probability Letters*, 82, 1337–1345.

Müller, H.G. (1991), "Smooth optimum kernel estimators near endpoints," *Biometrika*, 78, 521–530.

Müller, H.G., and Wang, J.L. (1994), "Hazard rate estimation under random censoring with varying kernels and bandwidths," *Biometrics*, pp. 61–76.

Scaillet, O. (2004), "Density estimation using inverse and reciprocal inverse Gaussian kernels," *Nonparametric Statistics*, 16, 217–226.

Schuster, E.F. (1985), "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistics-Theory and Methods*, 14, 1123–1136.

Wand, M.P., and Jones, M.C., *Kernel smoothing*, Vol. 60, Chapman & Hall/CRC (1994).