Asymptotic Results in Gamma Kernel Regression

Jianhong Shi[1] Weixing Song[2]

Shanxi Normal University, Linfen, Shanxi, China

and

Kansas State University, Manhattan, KS, USA

Abstract

Based on the Gamma kernel density estimation procedure, this paper constructs a nonparametric kernel estimate for the regression functions when the covariate are nonnegative. Asymptotic normality and uniform almost sure convergence results for the new estimator are systematically studied, and the finite performance of the proposed estimate is discussed via a simulation study and a comparison study with an existing method. Finally, the proposed estimation procedure is applied to the Geyser data set.

MSC: primary 62F35; secondary 62G10

Key words and phrases: Gamma Kernel Regression; Boundary Effect; Asymptotic Normality; Uniform Almost Sure Convergence

1 Introduction

Among all nonparametric procedures for estimating the density function of a random variable, the symmetric kernel density estimation method is the mostly commonly used one, see Rosenblatt (1956) and Parzen (1962) for an extensive introduction on this methodology. In principle, the symmetric kernel density estimation method can also be applied to the nonnegative data, but a major disadvantage of the symmetric kernel estimation technique is that the symmetric kernel function assigns positive weights outside the density support, thus induces the unpleasant boundary problem. This phenomenon is also carried over to the Nadaraya-Watson (N-W) estimators in regression setup. By data transformation, one can use the symmetric kernel methods to estimate the density functions of the nonnegative random variables, but statisticians are still interested in searching for methods to estimate the density function without the data transformation. More about the transformation methods,

[1]Jianhong Shi's research is supported by the Natural Science Foundation of Shanxi Province, China (2013011002-1).

[2]Corresponding author: weixing@ksu.edu. Weixing Song's research is partly supported by the NSF DMS 1205276.

see Wand, Marron and Ruppert (1991), Ruppert and Wand (1992), Marron and Ruppert (1994), among others.

Bagai and Rao (1996) proposed a nonnegative kernel method to estimate the density functions for nonnegative random variables, but their method has an undesirable feature that only the first several order statistics are used for estimating the density function. When density has a compact support, Chen (1999) proposed a Beta kernel density estimator, and when the data are supported on $[0, \infty)$, Chen (2000b) constructed a Gamma kernel density estimate. The biases and variances of these estimators are studied. 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. Scaillet (2004) proposed an inverse Gaussian kernel and a reciprocal inverse Gaussian kernel density estimate. Scaillet (2004) claimed that his procedure does not suffer from the boundary effect, but it is indeed not consistent at $x = 0$ when $f(0) > 0$. The varying kernel density estimate recently proposed by Mnatsakanov and Sarkisian (2012) is not consistent at $x = 0$ either. To remove the possible inconsistency or bias at $x = 0$, Chaubey et al. (2012) proposed a density estimator for non-negative random variables via smoothing of the empirical distribution function using a generalization of Hille's lemma. But the implementation of their method needs to select a perturbation parameter and a smoothing parameter.

The research is abundant on the asymptotic biases, variances, or mean square errors in the asymmetric kernel estimation. In comparison, the literature is rather scant on the investigation of the consistency and the asymptotic distributions of asymmetric kernel density and asymmetric kernel regression estimators. A generalized kernel smoothing technique is proposed in Chaubey, Laïb and Sen (2010) (CLS) to estimate the regression function in a class of non-negative stationary ergodic processes. By adding a perturbation term to a Gamma kernel, the resulting estimator successfully avoids the possible inconsistency at 0. Uniform consistency and asymptotic normality are also obtained, and simulation study shows that the proposed estimator is comparable to the local linear method. Similar to its counterpart in density estimation setup, to implement their estimation procedure, we have to select the perturbation and smoothing parameters. Some sporadic results can also be found in Bouezmarni and Rolin (2003), Chaubey et al. (2012), and Chen (2000a, 2002). In this paper, we will adopt the Gamma kernel proposed in Chen (2000b) to construct the estimator. Similar to the method in CLS, the proposed estimator does not have the inconsistency problem at 0, but only one smoothing parameter is needed. Although we only discuss the large sample behavior of the proposed estimator in the i.i.d. context, the extension to the non-negative stationary ergodic processes is not difficult.

The paper is organized as follows. Section 2 introduces the Gamma kernel density and regression estimation procedures. Technical assumptions and the large sample results of the proposed estimates will be present in Section 3. Some further discussion on the selection of bandwidth will be also made. Findings of a simulation study and a comparison study, together with a real data example, are presented in Section 4, and the proofs of the main results appear in Section 5.

Unless specified otherwise, all limits are taken as $n \to \infty$, and c will denote a generic positive constant whose value depends on the context.

2 Gamma Kernel Regression Estimation

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a population X supported on $[0, \infty)$. Let $K_{p,q}$ be the density function a Gamma(p, q) random variable. For any fixed $x \geq 0$, the first class of Gamma kernel estimator for the density function f of X proposed by Chen (2000b) is defined as

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x/h_n+1,h_n}(X_i) = \frac{\sum_{i=1}^{n} X_i^{x/h_n} \exp(-X_i/h_n)}{n\Gamma(x/h_n + 1) h_n^{x/h_n+1}}, \tag{2.1}$$

where h_n is a sequence smoothing parameters satisfying the conditions that $h_n \to 0$, $nh_n \to \infty$ as $n \to \infty$. Different from the Gamma kernel density used in CLS, $\hat{f}_n(s)$ defined in (2.1) has definition at $x = 0$, and this is the very feature that makes our regression estimator free from the possible inconsistency at 0. To achieve the same objective, CLS adds a perturbation term to the x-values where the regression function is evaluated, which noticeably stretches the computational time when searching for the optimal smoothing parameters. For the sake of simplicity, the subscript n will be suppressed from h_n in the sequel. Chen (2000b) showed that $E\hat{f}_n(x) = f(x) + h[f'(x) + xf''(x)/2] + o(h)$. Here and in the sequel, for any function $g(x)$, $g'(x)$, $g''(x)$ denote the first and second derivatives of $g(x)$ with respect to x, respectively. So the bias of $\hat{f}_n(x)$ is $O(h)$ for all $x \in [0, \infty)$, the density estimator does not suffer from the boundary bias. A less desirable feature of $\hat{f}_n(x)$ is the presence of f' in the bias, which could be removed by replacing $K_{x/h+1,h}$ with $K_{\rho_h(x),h}$ in (2.1), where

$$\rho_h(x) = \frac{x}{h} I(x \geq 2h) + \left[\frac{x^2}{4h^2} + 1\right] I(0 \leq x < 2h),$$

and $I(\cdot)$ is the indicator function. More properties of $\hat{f}_n(x)$ and its modified version, including the expressions of their variances, mean squared errors (MSE) and mean integrated squared errors, can also be found in Chen (2000b). In this paper, we will find the asymptotic distribution of $\hat{f}_n$, and extend Chen (2000b)'s idea to the nonparametric regression setup based on the kernel $K_{x/h+1,h}$. Although the involvement of f' in the bias will be carried over to the regression estimation, the easiness of derivation of theoretical properties and the concise presentation of the asymptotic results would make this sacrifice worthwhile. Most importantly, the arguments we developed for the regression estimator based on $K_{x/h+1,h}$ can be easily adapted to the one constructed from the modified kernel $K_{\rho_h(x),h}$.

There is a very interesting connection between the Gamma kernel in Chen (2000b) and the normal kernel used in the symmetric kernel density estimation. For fixed $x > 0$, let T_h denote the random variable following a Gamma$(x/h + 1, h)$ distribution. It can be verified that $(hx)^{-1/2}(T_h - x) \Longrightarrow_d N(0, 1)$ as $h \to 0$. This implies that asymptotically, as $h \to 0$, the Gamma kernel behaves like the normal kernel in which different bandwidths are used for each point x at which $f(x)$ is estimated. This is so called local kernel density estimator, see Schucany (1989), Wand and Jones (1995). Similar phenomenon can be found for the varying kernel by Mnatsakanov and Sarkisian (2012).

The relationship between a scalar response Y and a one-dimensional covariate X is often investigated through the regression model $Y = m(X) + \varepsilon$, where ε accounts for the random error with usual assumptions $E(\varepsilon|X = x) = 0$ and $\sigma^2(x) := E(\varepsilon^2|X = x) > 0$, for almost all x. Analogous to the N-W kernel regression estimate, the Gamma kernel regression estimate of $m(x)$ when the covariate X is nonnegative is defined as

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n K_{x/h+1,h}(X_i)Y_i}{\sum_{i=1}^n K_{x/h+1,h}(X_i)}. \tag{2.2}$$

Under some regularity conditions on the underlying density function $f(x)$ and the regression function $m(x)$, asymptotic normality of the Gamma kernel estimate $\hat{m}_n(x)$, as well as its uniform consistency, is established in the paper.

A simpler expression for $\hat{m}_n(x)$ in (2.2) can be derived from the definition of $K_{x/h+1,h}$. In fact, after some cancelation, we have

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n X_i^{x/h} \exp(-X_i/h)Y_i}{\sum_{i=1}^n X_i^{x/h} \exp(-X_i/h)}. \tag{2.3}$$

This formula is mainly useful for the computation of $\hat{m}_n$ while (2.2) is convenient for theoretical development. Similar to the symmetric kernel regression, to deal with the multidimensional covariate cases, one can use the product kernel in the definition of the regression function estimate. For the sake of brevity, we will not pursue this topic here.

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

In this section, we shall present the asymptotic expressions of the conditional bias, variance, hence the MSE of $\hat{m}_n(x)$ defined in (2.2), and the asymptotic normality of $\hat{f}_n(x)$ and $\hat{m}_n(x)$. 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 Gamma kernel function. We start with some technical assumptions needed for the asymptotic theories.

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

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

(A2). $E(\varepsilon|X) = 0$, and the second order derivatives of fm, fm^2 are 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 $x > 0$, and the second order derivative of $f\sigma^2$ is continuous and bounded on $[0, \infty)$.

(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). $h \to 0$, $n\sqrt{h} \to \infty$ as $n \to \infty$.

Condition (A1) on $f(x)$ is also adopted by Chen (2000b) implicitly when deriving MSE of $\hat{f}_n(x)$. Similar to (A1), Condition (A2) is needed to control the higher order term in the asymptotic expansions of MSE for $\hat{m}_n(x)$. 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), similar to its symmetric kernel context, is a minimal condition needed for the smoothing parameter. Additional assumptions on h as needed are stated in various theorems presented below.

3.1 Conditional Bias and Variance

Define

$$b(x) = m'(x) + \frac{1}{2}xm''(x) + \frac{xm'(x)f'(x)}{f(x)}, \quad v(x) = \frac{\sigma^2(x)}{2f(x)\sqrt{\pi x}}, \tag{3.1}$$

and $\mathbf{X} := \{X_1, X_2, \ldots, X_n\}$. The following theorem presents the conditional biases and variances of $\hat{m}_n(x)$ defined in (2.2).

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

(i). For $x > 0$,

$$\text{Bias}(\hat{m}_n(x)|\mathbf{X}) = hb(x) + o_p(h) + O_p(h^{1/4}/\sqrt{n}), \tag{3.2}$$

$$\text{Var}(\hat{m}_n(x)|\mathbf{X}) = \frac{v(x)}{n\sqrt{h}} + o_p\Big(\frac{1}{n\sqrt{h}}\Big). \tag{3.3}$$

(ii). For $x = 0$,

$$\text{Bias}(\hat{m}_n(0)|\mathbf{X}) = hm'(0) + o_p(h), \tag{3.4}$$

$$\text{Var}(\hat{m}_n(0)|\mathbf{X}) = \frac{\sigma^2(0)}{2nhf(0)} + o_p\Big(\frac{1}{nh}\Big). \tag{3.5}$$

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

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

when $x > 0$, and

$$MSE(\hat{m}_n(0)|\mathbf{X}) = h^2 m'^2(0) + \frac{\sigma^2(0)}{2nhf(0)} + o_p(h^2) + o_p\left(\frac{1}{nh}\right)$$

when $x = 0$.

Theorem 3.1 indicates that the conditional biases of $\hat{m}_n(x)$ have the same order at interior points and at 0, but the conditional variance has a larger order at 0 than at interior points.

Similar to the N-W kernel regression case, one can choose the optimal smoothing parameter h by minimizing the leading term in the conditional or unconditional MSE of $\hat{m}_n$ with respect to h. For $x > 0$, we can verify that h 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. However, for $x = 0$, h has the order of $n^{-1/3}$, and the corresponding MSE has the order of $n^{-2/3}$, slower than the convergence rate of MSE when x is an interior point.

3.2 Asymptotic Normality

Although the asymptotic normality of $\hat{f}_n(x)$ is not discussed in Chen (2000b), the derivation of such result should be straightforward based on the asymptotic expansions for the bias, variance obtained in the paper. For the sake of completeness, we shall present the asymptotic normality result for $\hat{f}_n(x)$ along with its proof in Section 5. As its further development in the regression setup, the asymptotic normality of $\hat{m}_n(x)$ will be also reported in this section.

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

$$\Big(\frac{f(x)}{2n\sqrt{\pi x h}}\Big)^{-1/2}\Big[\hat{f}_n(x) - f(x) - h[f'(x) + xf''(x)/2] + o(h)\Big] \to_d N(0,1). \tag{3.6}$$

For $x = 0$, $f(0) > 0$, if further assume that $nh \to \infty$, then

$$\Big(\frac{f(0)}{2nh}\Big)^{-1/2}\Big[\hat{f}_n(0) - f(0) - hf'(0) + o(h)\Big] \to_d N(0,1). \tag{3.7}$$

For $x = 0$, $f(0) = 0$, $f'(0) \neq 0$, if further assume that $nh^2 \to \infty$, then

$$\Big(\frac{f'(0)}{4n}\Big)^{-1/2}\Big[\hat{f}_n(0) - hf'(0) + o(h)\Big] \to_d N(0,1). \tag{3.8}$$

It should be noted that the last result only has a theoretical value.

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 $f(x) > 0$,*

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

where $b(x)$ and $v(x)$ are defined in (3.1). For $x = 0$, but $f(0) > 0$,

$$\Big(\frac{\sigma^2(0)}{2nhf(0)}\Big)^{-1/2}\Big[\hat{m}_n(0) - m(0) - hm'(0) + o_p(h)\Big] \to_d N(0,1).$$

If we further assume that $\log n/(n\sqrt{h}) \to 0$, then $o_p(1)$ can be replaced by $o(1)$ in the above results.

It is noted that there is a non-negligible asymptotic bias appearing in the above results, a characteristic shared with the N-W kernel density and regression estimates. These biases can be eliminated by under-smoothing which, in the current set up, is to select a proper h such that $nh^{5/2} \to 0$ for $x > 0$ and $nh^3 \to 0$ for $x = 0$, without violating conditions $h \to 0, n\sqrt{h} \to \infty$. The large sample confidence intervals for $m(x)$ thus can be constructed with the help of Theorem 3.3.

The superiority of the proposed estimator to the CLS estimator can be easily seen by comparing the variances of the corresponding limiting normal distributions. For the interior points, both estimators have a convergence rate $n^{-2/5}$, but the asymptotic variance of CLS estimator is $\sigma^2(x)/(2\sqrt{\pi}xf(x))$, and that of the proposed estimator is $\sigma^2(x)/(2\sqrt{\pi x}f(x))$. Then for x values close to 0, our estimator has smaller variance than CLS estimator. For the boundary point $x = 0$, the optimal MSE convergence rate of the proposed estimator is $n^{-2/3}$, while CLS does not have a universal optimal convergence rate. Based on one choice of the smoothing parameters, the convergence rate for CLS estimator is $n^{-4/7}$, which is slower than $n^{-2/3}$. But the asymptotic variance of CLS estimator at $x = 0$ is $\sigma^2(0)/(2\sqrt{\pi}f(x))$ which is smaller than the asymptotic variance of the proposed estimator, $\sigma^2(0)/(2f(0))$.

3.3 Uniform Almost Sure Consistency

In this section we develop an almost sure uniform convergence result for $\hat{m}_n(x)$ over an arbitrary closed sub-interval of $(0,\infty)$. To do this we apply Borel-Cantelli lemma and the Bernstein inequality, after verifying the Cramér condition: for some $k \geq 2$, $c > 0$, and h small enough,

$$E|K_{x/h+1,h}(X)|^k \leq k!\Big(\frac{c}{n\sqrt{h}}\Big)^{k-2} EK^2_{x/h+1,h}(X), \quad x > 0. \tag{3.9}$$

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* $\log n/n\sqrt{h} \to 0$. *Then for any constants* a *and* b *such that* $0 < a < b < \infty$,

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

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

$$\sup_{x\in[a,b]} \Big|\hat{m}_n(x) - m(x)\Big| = O(h) + o\Big(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\Big), \quad a.s.$$

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.

Remark: Similar to the N-W kernel regression, the unconditional version of Theorem 3.1 is very hard to derive. Using Fan (1993)'s argument, we can show that the unconditional version of Theorem 3.1 remains valid for

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

The proof of this claim follows the same argument as the proof of Theorem 3 in Fan (1993), hence is omitted here for the sake of brevity.

Another important issue is the selection of the smoothing parameter h, which affects greatly the finite sample performance of the Gamma kernel regression estimate. It is desirable to construct some practical data-driven selection procedures. As in the symmetric kernel regression cases, we can develop some procedures, such as the least square cross validation

(LSCV) or k-fold LSCV, as well as the generalized cross validation (GCV) procedure. For example, the GCV procedure from the N-W kernel regression can be adapted to the current setup as follows. Define

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

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

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

Some other smoothing parameter selection procedures can also be considered, such as AIC or BIC-type criteria.

4 Numerical Study

To evaluate the finite sample performance of the proposed Gamma kernel regression estimate, a simulation study and a comparison study with CLS method, and a real data analysis are conducted in this section.

Simulation Study: In the simulation, the underlying density function of the design variable is chosen to be log-normal with $\mu = 0, \sigma = 1$ with scale parameter 1, and the random error ε to be normal with mean 0 and standard deviation 0.5. Three simple regression functions, $m(x) = 1/x^2$, $m(x) = (x-1)^2$, and $m(x) = x^2$ are considered. For $m(x) = 1/x^2$, the estimate will be evaluated at 500 equally spaced values over the interval $(0.2, 2)$; for $m(x) = (x-1)^2$ and $m(x) = x^2$, the estimate will be evaluated at 500 equally spaced values over the interval $(0, 2)$, and the sample sizes used are 100, 200 and 300. Then the MSEs between the estimated values and true values of the regression function will be used for comparison. To see the effect of bandwidth on the finite sample performance of the Gamma kernel regression estimate, we choose the optimal bandwidth based on MSE criterion stated in Section 2, that is, $h = an^{-2/5}$, where a is chosen to be $0.2, 0.4, 0.6, 0.8$.

Table 1 presents the MSEs from the simulation study. Clearly, the estimation of $1/x^2$ does not perform very well. However, for $m(x) = (x-1)^2$ or x^2, the MSEs are relatively stable, and reasonably small across all the choices of a values. Also it seems like that a smaller bandwidth produces a better fit.

Comparison Study: We now conduct a simulation study to compare the finite sample performance of the proposed method with the CLS estimator, denoted by $\tilde{m}_n(x)$. Let ε_n and

	$m(x)$	$1/x^2$	$(x-1)^2$	x^2	$m(x)$	$1/x^2$	$(x-1)^2$	x^2
n=100	$a=0.2$	0.2677	0.0098	0.0112	$a=0.6$	1.2543	0.0045	0.0181
	$a=0.4$	0.6479	0.0048	0.0122	$a=0.8$	1.9254	0.0067	0.0237
n=200	$a=0.2$	0.0586	0.0095	0.0072	$a=0.6$	0.2154	0.0049	0.0055
	$a=0.4$	0.0649	0.0063	0.0045	$a=0.8$	0.4891	0.0051	0.0077
n=300	$a=0.2$	0.3013	0.0026	0.0274	$a=0.6$	0.8098	0.0067	0.0405
	$a=0.4$	0.6166	0.0035	0.0368	$a=0.8$	0.9831	0.0110	0.0423

Table 1: Mean Squared Errors of Regression Estimates

v_n be two smoothing parameters. The estimator $\tilde{m}_n(x)$ has the following form

$$\tilde{m}_n(x) = \frac{\sum_{i=1}^n Q_{x+\varepsilon_n,v_n}(X_i)Y_i}{\sum_{i=1}^n Q_{x+\varepsilon_n,v_n}(X_i)},$$

where

$$Q_{x+\varepsilon_n,v_n}(t) = \frac{t^{1/v_n^2-1}}{[v_n^2(x+\varepsilon_n)]^{1/v_n^2}\Gamma(1/v_n^2)} \exp\left(-\frac{t}{v_n^2(x+\varepsilon_n)}\right).$$

A simpler form of $\tilde{m}_n(x)$ is given by

$$\tilde{m}_n(x) = \frac{\sum_{i=1}^n X_i^{1/v_n^2-1} \exp\left(-\frac{X_i}{v_n^2(x+\varepsilon_n)}\right) Y_i}{\sum_{i=1}^n X_i^{1/v_n^2-1} \exp\left(-\frac{X_i}{v_n^2(x+\varepsilon_n)}\right)}.$$

Comparing to the proposed estimator $\hat{m}_n(x)$ defined in (2.3), the x argument are placed in the different locations. As a result, small x values might bring some computational instability to $\tilde{m}_n(x)$, but not to $\hat{m}_n(x)$. Also, two smoothing parameters are needed in $\tilde{m}_n(x)$, but only one is needed in $\hat{m}_n(x)$. Therefore, when searching for optimal smoothing parameter values, our method will be computationally more efficient than $\tilde{m}_n(x)$.

We will adopt similar models as the one used in CLS for the comparison study. Two regression models are considered, $m_1(x) = x + 2\exp(-16x^2)$ and $m_2(x) = \sin(2x) + 2\exp(-16x^2)$, X will be taken from the exponential distribution with rate parameter 1. As for the random error, we will consider two cases, a normal random error and a double exponential (DE) error, both with mean 0 and variance 0.5^2.

Leave-one-out CV procedure will be used to search the optimal smoothing parameters for both estimators. That is, the minimizers of $R_n = n^{-1}\sum_{i=1}^n [Y_i - M_{n,(-i)}(X_i)]^2$ will be the smoothing values, where $M_{n,(-i)}(X_i)$ stands for either $\hat{m}_n$ or $\tilde{m}_n$, evaluated at X_i with (X_i, Y_i) removed from the sample. To speed up the computation process, for CLS estimator, we choose $\varepsilon_n = v_n^2$, and for both estimators, we search the optimal h and v_n values from 200 equally spaced grid points from $[0.001, 1]$.

The sample sizes are selected to be $100, 150$ and 200. For each sample size, we replicate the estimation procedures 200 times, and the entries in Table 2 report the average values of the minimum R_n for each estimator. In each cell, the first number corresponds to our proposed estimator, and the second one is for the CLS estimator.

	$m_1(x)$		$m_2(x)$	
n	$\varepsilon \sim N(0, 0.5^2)$	$\varepsilon \sim DE(0, 0.5^2)$	$\varepsilon \sim N(0, 0.5^2)$	$\varepsilon \sim DE(0, 0.5^2)$
100	0.3023655	0.2960982	0.2907079	0.2803470
	0.3123051	0.3042341	0.2998222	0.2896394
150	0.2899166	0.2798531	0.2775187	0.2832459
	0.2966342	0.2853773	0.2867333	0.2920715
200	0.2704181	0.2730069	0.2717700	0.2808886
	0.2765468	0.2782480	0.2801365	0.2902728

Table 2: Average of the minimum R_n from 200 replications

From the simulation result, it is interesting to notice that for both regression models, these two estimators perform almost equally well, with a little bit advantage for our method over the CLS method. These observations, together with the relative easiness in finding the smoothing parameter, make our estimator a very strong, if not stronger, competitor to CLS method, which, in turn, is a good competitor to the well established local linear procedure.

Real Data Example: Finally, We apply the Gamma kernel regression estimation procedure for the data set from Azzalini and Bowman (1990) on the Old Faithful geyser in Yellowstone National Park, Wyoming, USA. The data consist of 299 pairs of measurements on the waiting time X to the next eruption, and the eruption time Y in minutes, which were collected continuously from August 1st until August 15th, 1985. We shall use the nonparametric regression procedure developed in this paper to investigate the relationship between Y and X. To prevent the exponential operation in the computation of the Gamma kernel regression estimate from explosion, the waiting time X will be transformed to $T = (X - 43)/30$ first. As a result, the range of T is between 0 and 2.17. Here 43 is the minimum of X.

The scatter plot in Figure 1 shows the data structure with Y against T. The bandwidth h is chosen based on the leave-one-out LSCV criterion, which is $h = 0.0110$. The regression curve is imposed on the scatter plot in Figure 1 with solid line. For comparison purpose, the N-W kernel regression curve is also drawn in Figure 1 but with dashed line. Again, the leave-one-out LSCV criterion is applied to select the bandwidth, which gives $h = 0.1215$. Like the N-W kernel regression estimate, the Gamma kernel regression estimate also captures the main characteristic of the data structure, but it appears less variable than the former.

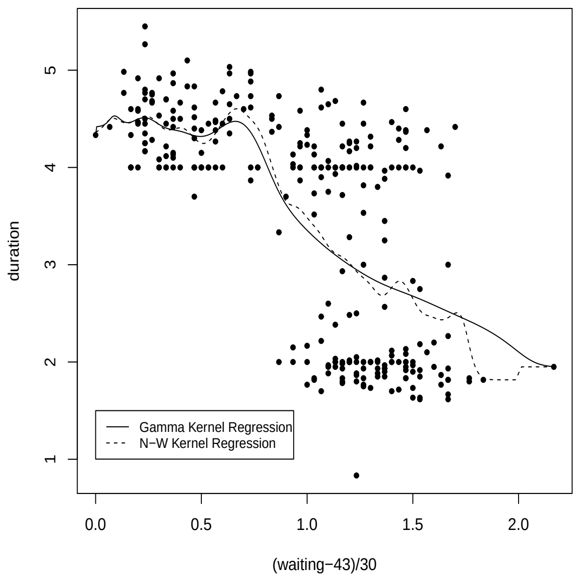

Figure 1: Regression of Geyser Data

5 Proofs of the Main Results

This section contains the proofs of all the large sample results presented in Section 3. 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 an Gamma distribution with shape parameter p and scale parameter λ is

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

Its mean μ and variance τ^2 are $\mu = \lambda p$, and $\tau^2 = \lambda^2 p$. For any $k \geq 1$, Let $p_k = kx/h+1, \lambda_k = h/k, k = 1, 2, \cdots, x > 0$. Write μ_k, τ_k, for μ, τ, 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 5.1 *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 $x \geq 0$ and $k \geq 1$,*

$$\int_0^\infty g(u; p_k, \lambda_k) l(u) du = l(x) + \frac{[2l'(x) + xl''(x)]h}{2k} + o(h). \tag{5.1}$$

Proof of Lemma 5.1. Fix an $x \geq 0$. Note that $\mu_k = \lambda_k p_k = x + h/k$. Taylor expansion of $l(\mu_k)$ around x up to the second order yields

$$l(\mu_k) = l(x) + \frac{hl'(x)}{k} + \frac{h^2 l''(\tilde{\xi})}{2k^2}, \tag{5.2}$$

where $\tilde{\xi}$ is some value between μ_k and x.

Taylor expansion of $l(u)$ at the mean μ_k of $g(u;p_k,\lambda_k)$ gives

$$\int_0^\infty g(u,p_k,\lambda_k)l(u)du = l(\mu_k) + \frac{1}{2}\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)l''(\tilde{u})du, \tag{5.3}$$

where, for any $u \in [0,\infty)$, $\tilde{u}$ is some value between u and μ_k.

For $x = 0$, $\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)du = \tau_k^2 = h^2/k^2$, so by the boundedness of l'', we have

$$\left|\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)l''(\tilde{u})du\right| \le c\tau_k^2 = o(h) \tag{5.4}$$

for some positive constant c. This, together with (5.2) implies that (5.1) holds for $x = 0$.

Now let's assume $x > 0$. Rewrite (5.3) as

$$\int_0^\infty g(u,p_k,\lambda_k)l(u)du = l(\mu_k) + \frac{1}{2}l''(\mu_k)\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)du$$
$$+\frac{1}{2}\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du. \tag{5.5}$$

Note that the first integration on the right hand side of (5.5) is the variance of $g(u,p_k,\lambda_k)$, $\tau_k^2 = xh/k + h^2/k^2$. From (5.2) and the continuity of l'', we can verify that the two leading terms on the right hand side of (5.5) match the expansion (5.1). So it is sufficient to show that the third term on the right hand side of (5.5) is of the order $o(h)$.

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

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 gamma density function $g(u,p_k,\lambda_k)$ is unimodal, and the mode is $\lambda_k(p_k - 1) = x$. Therefore, $g(u,p_k,\lambda_k) \le g(\delta_1,p_k,\lambda_k)$ for all $0 < u < \delta_1$, and

$$\int_0^{\delta_1} (u-\mu_k)^2 g(u;p_k,\lambda_k)du \le g(\delta_1,p_k,\lambda_k)\int_0^{\delta_1} (u-x-h/k)^2 du = g(\delta_1,p_k,\lambda_k)O(1).$$

For h small enough, by Stirling approximation,

$$\frac{1}{\Gamma(p_k)\lambda_k^{p_k}} = \frac{1}{\Gamma(kx/h+1)\cdot(h/k)^{kx/h+1}} = \frac{\sqrt{k}\exp(kx/h)}{x^{kx/h+1/2}\sqrt{2\pi h}}[1+o(1)]. \tag{5.6}$$

Then we have

$$\begin{aligned} g(\delta_1,p_k,\lambda_k) &= \frac{\sqrt{k}\exp(kx/h)}{x^{kx/h+1/2}\sqrt{2\pi h}}\delta_1^{kx/h}\exp(-\delta_1 k/h)[1+o(1)] \\ &= \sqrt{\frac{1}{2\pi hx}}\left(\frac{\delta_1}{x}\exp\left(1-\frac{\delta_1}{x}\right)\right)^{kx/h}[1+o(1)] = o(h), \end{aligned}$$

because $0 < \delta_1/x < 1$, and the fact that as a function of x, $x\exp(1-x) < 1$ for all $x > 0$ and $x \neq 1$. This implies

$$\left|\int_0^{\delta_1} (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du\right| = o(h). \tag{5.7}$$

Let $\delta_2 = x + \gamma_2$. The boundedness of l'' implies

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

But,

$$\int_{\delta_2}^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)du = \frac{1}{\Gamma(p_k)\lambda_k^{p_k}}\int_{\delta_2}^{\infty} (u-\mu_k)^2 u^{p_k-1}\exp\left(-\frac{u}{\lambda_k}\right)du,$$

while the right hand side is bounded above by

$$\frac{4}{\Gamma(p_k)\lambda_k^{p_k}}\int_{\delta_2}^{\infty} u^{-2}u^{p_k+3}\exp\left(-\frac{u}{\lambda_k}\right)du.$$

As a function of u, $u^{p_k+3}\exp(-u/\lambda_k)$ is increasing on $(0, \lambda_k(p_k+3)) = (0, x+3h/k)$, and decreasing on $(\lambda_k(p_k+3), \infty) = (x+3h/k, \infty)$. For h small enough, we have $x + 3h/k \leq \delta_2$. So for any $u \geq \delta_2$, $u^{p_k+3}\exp(-u/\lambda_k) \leq \delta_2^{p_k+3}\exp(-\delta_2/\lambda_k)$. Thus, by (5.6),

$$\begin{aligned}\int_{\delta_2}^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)du &\leq \frac{\sqrt{k}\exp(kx/h)}{x^{kx/h+1/2}\sqrt{2\pi h}} \cdot \delta_2^{kx/h+4}\exp\left(-\delta_2 k/h\right)\int_{\delta_2}^{\infty} u^{-2}du[1+o(1)] \\ &= O(h^{-1/2}) \cdot \left(\frac{\delta_2}{x}\exp\left(1-\frac{\delta_2}{x}\right)\right)^{kx/h} = o(h)\end{aligned}$$

again by $(\delta_2/x)\exp(1-\delta_2/h) < 1$ for $\delta_2 > x$. Therefore,

$$\left|\int_{\delta_2}^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du\right| = o(h). \tag{5.8}$$

Finally, we shall show that

$$\int_{\delta_1}^{\delta_2} (u-\mu_k)^2 g(u; p_k, \lambda_k)[l''(\tilde{u}) - l''(\mu_k)]du = o(h). \tag{5.9}$$

By uniform continuity of l'',

$$\int_{\delta_1}^{\delta_2} (u-\mu_k)^2 g(u; p_k, \lambda_k)|l''(\tilde{u}) - l''(\mu_k)|du \leq \varepsilon\int_0^{\infty} (u-\mu_k)^2 g(u; p_k, \lambda_k)du,$$

by the fact that $|\tilde{u} - \mu_k| \leq |u - \mu_k| < \gamma$, for $u \in [\delta_1, \delta_2]$ and h sufficiently small. Because $\int_0^{\infty}(u-\mu_k)^2 g(u; p_k, \lambda_k)du = \tau_k^2 = O(h)$, we obtain

$$\int_{\delta_1}^{\delta_2} (u-\mu_k)^2 g(u; p_k, \lambda_k)|l''(\tilde{u}) - l''(\mu_k)|du = \varepsilon \cdot O(h). \tag{5.10}$$

The arbitrariness of ε, combined with (5.7), (5.8) and (5.10), finally yield

$$\int_0^\infty (u-\mu_k)^2 g(u;p_k,\lambda_k)[l''(\tilde{u})-l''(\mu_k)]du = o(h)$$

for $x>0$. This, together with (5.4), completes the proof of the Lemma 5.1. □

The same conclusion can be obtained with a simpler proof by assuming that the second derivative of l'' is Hölder continuous.

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

$$\hat{m}_n(x)-m(x) = \frac{B_n(x)+V_n(x)}{f(x)} + \Big[\frac{1}{\hat{f}_n(x)} - \frac{1}{f(x)}\Big][B_n(x)+V_n(x)], \tag{5.11}$$

where

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

with $K_{x/h+1,h}(X_i)$ defined in (2.1). Now we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. From the assumption (A2), the conditional bias of $\hat{m}_n(x)$ equals $E[\hat{m}_n(x)|\mathbf{X}]-m(x) = B_n(x)/\hat{f}_n(x)$. From Chen (2002), it is known that $\hat{f}_n(x) = f(x)+o_p(1)$. In the following, we shall develop asymptotic expansions of the expectation and variance of $B_n(x)$.

First, let's consider the expectation of $B_n(x)$ which can be written as

$$EB_n(x) = EK_{x/h+1,h}(X)m(X) - m(x)EK_{x/h+1,h}(X).$$

Applying Lemma 5.1 with $l = H = mf$ and $l = f$, $k=1$, by assumption (A1) and (A2),

$$EK_{x/h+1,h}(X)m(X) = H(x) + \frac{[2H'(x)+xH''(x)]h}{2} + o(h),$$
$$m(x)EK_{x/h+1,h}(X) = m(x)\left[f(x) + \frac{[2f'(x)+xf''(x)]h}{2} + o(h)\right].$$

Therefore,

$$EB_n(x) = \left[m'(x)f(x) + \frac{x}{2}m''(x)f(x) + xm'(x)f'(x)\right]h + o(h) \tag{5.12}$$

Direct calculation shows that $m'(x)f(x)+xm''(x)f(x)/2+xm'(x)f'(x) = b(x)f(x)$, where $b(x)$ is defined in (3.1).

Now consider the variance of $B_n(x)$. Since the variance of a random variable is bounded above by its second moment, so

$$\mathrm{Var}(B_n(x)) \le \frac{1}{n}E\left[\frac{1}{h\Gamma(x/h+1)}\left(\frac{X}{h}\right)^{x/h}\exp\left(-\frac{X}{h}\right)[m(X)-m(x)]\right]^2 \tag{5.13}$$

which can be written as

$$\begin{aligned}
&\frac{1}{n}E\left[\frac{1}{h\Gamma(x/h+1)}\left(\frac{X}{h}\right)^{x/h}\exp\left(-\frac{X}{h}\right)[m(X)-m(x)]\right]^2\\
&=\frac{1}{n}\int_0^\infty \frac{1}{h^2\Gamma^2(x/h+1)}\left(\frac{u}{h}\right)^{2x/h}\exp\left(-\frac{2u}{h}\right)[m(u)-m(x)]^2 f(u)du\\
&=\frac{\Gamma(2x/h+1)}{nh2^{2x/h+1}\Gamma^2(x/h+1)}\int_0^\infty K_{p_2,\lambda_2}(u)f(u)[m(u)-m(x)]^2du.
\end{aligned}$$

For $x > 0$, by Stirling approximation, as $h \to 0$, we get

$$\frac{\Gamma(2x/h+1)}{nh2^{2x/h+1}\Gamma^2(x/h+1)}=\frac{1}{2n\sqrt{\pi x h}}[1+o(1)]. \tag{5.14}$$

Then by the continuity of f, m, f', m' and (A2), one can show that

$$\int_0^\infty K_{p_2,\lambda_2}(u)f(u)[m(u)-m(x)]^2du=\frac{xf(x)m'^2(x)h}{2}+o(h). \tag{5.15}$$

Combining (5.14) and (5.15), we get $\text{Var}(B_n(x)) = O(\sqrt{h}/n)$, which, together with (5.12), implies (3.2). For $x = 0$, by (5.13),

$$\begin{aligned}
\text{Var}(B_n(0)) &\leq \frac{1}{n}E\left[\frac{1}{h}\exp\left(-\frac{X}{h}\right)[m(X)-m(0)]\right]^2\\
&=\frac{1}{nh}\int_0^\infty \exp\left(-\frac{2u}{h}\right)[m(u)-m(0)]^2 f(u)du.
\end{aligned}\tag{5.16}$$

Again by the continuity of f, m, f', m' and (A2), one can show that

$$\int_0^\infty \exp\left(-\frac{2u}{h}\right)[m(u)-m(0)]^2 f(u)du = O(h^2). \tag{5.17}$$

Combining (5.16) and (5.17), we get $\text{Var}(B_n(0)) = O(h/n)$, which, together with (5.12), implies (3.4).

Next, consider the conditional variance $\text{Var}[\hat{m}_n(x)|\mathbf{X}]$. It is easily seen that

$$\text{Var}[\hat{m}_n(x)|\mathbf{X}]=\frac{1}{n^2h^2\Gamma^2(x/h+1)\hat{f}_n^2(x)}\sum_{i=1}^n\left(\frac{X_i}{h}\right)^{2x/h}\exp\left(-\frac{2X_i}{h}\right)\sigma^2(X_i).$$

Applying Lemma 5.1 with $l = f\sigma^2$, by (A3), one can show that, for $x > 0$

$$\frac{1}{nh^2\Gamma^2(x/h+1)}E\left(\frac{X}{h}\right)^{2x/h}\exp\left(-\frac{2X}{h}\right)\sigma^2(X)=\frac{\sigma^2(x)f(x)}{2n\sqrt{\pi x h}}+O\left(\frac{\sqrt{h}}{n}\right). \tag{5.18}$$

Therefore, from $\hat{f}_n(x) = f(x) + o_p(1)$, one can obtain (3.3). If $x = 0$, (5.18) becomes

$$\frac{1}{nh^2}E\exp\left(-\frac{2X}{h}\right)\sigma^2(X)=\frac{\sigma^2(0)f(0)}{2nh}+O\left(\frac{1}{n}\right).$$

Hence, from $\hat{f}_n(0) = f(0) + o_p(1)$, one can obtain (3.5). $\square$

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

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

For any $x \geq 0$, from Chen (2000b), or a direct application of Lemma 5.1 with $l \equiv 1$, we have $EK_{x/h+1,h}(X) = f(x) + [f'(x) + xf''(x)/2]h + o(h)$, therefore,

$$\hat{f}_n(x) - f(x) - \left[f'(x) + \frac{xf''(x)}{2}\right] h + o(h) = \sum_{i=1}^{n} \xi_{in}(X_i). \tag{5.19}$$

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

$$E|\xi_{in}(x)|^{2+\delta} \leq n^{-(2+\delta)} 2^{1+\delta} [EK^{2+\delta}_{x/h+1,h}(X) + (EK_{x/h+1,h}(X))^{2+\delta}].$$

Let $p_\delta = (2+\delta)x/h + 1$, $\lambda_\delta = h/(2+\delta)$. We can show that

$$EK^{2+\delta}_{x/h+1,h}(X) = B_h(x) \int_0^\infty g(u, p_\delta, \lambda_\delta) f(u) du,$$

where

$$B_h(x) = \frac{\Gamma((2+\delta)x/h + 1)}{(2+\delta)^{(2+\delta)x/h+1} h^{1+\delta} \Gamma^{2+\delta}(x/h+1)}.$$

Applying Stirling approximation to the Gamma functions in $B_h(x)$, we have, for h sufficiently small and $x > 0$,

$$B_h(x) = \frac{1}{(2\pi x h)^{(1+\delta)/2} \sqrt{2+\delta}} [1 + o(1)],$$

which, together with the result $\int_0^\infty g(u, p_\delta, \lambda_\delta) f(u) du = f(x) + (1)$, implies $EK^{2+\delta}_{x/h+1,h}(X) = O(h^{-(1+\delta)/2})$. From Chen (2000b), we also know that

$$v_n^2(x) = \mathrm{Var}\left[\sum_{i=1}^{n} \xi_{in}(x)\right] = \mathrm{Var}(\hat{f}_n(x)) = \frac{f(x)}{2n\sqrt{\pi h x}} + o(1/n\sqrt{h}).$$

This, together with $EK_{x/h+1,h}(X) = f(x) + o(1)$, implies

$$\begin{aligned} v_n^{-(2+\delta)}(x) \sum_{i=1}^{n} E\xi_{in}^{2+\delta}(x) &= n v_n^{-(2+\delta)}(x) E\xi_{1n}^{2+\delta}(x) \\ &= n \cdot n^{-(2+\delta)} \cdot O\left((n\sqrt{h})^{(2+\delta)/2}\right) \cdot O(h^{-(1+\delta)/2}) \\ &= O((n\sqrt{h})^{-\delta/2}) = o(1) \end{aligned}$$

by (A5). Therefore, by Lindeberg-Feller central limit theorem, $v_n^{-1}(x)\sum_{i=1}^n \xi_{in}(x) \Rightarrow_d N(0,1)$. By (5.19), (3.6) is proved.

For $x=0$ and $f(0)>0$, from Chen (2000b) or Lemma 5.1, one can show that

$$E\hat{f}_n(0) = f(0) + f'(0)h + o(h), \quad \mathrm{Var}(\hat{f}_n(0)) = \frac{f(0)}{2nh} + o((nh)^{-1}).$$

For any $\delta > 0$, we still have

$$E|\xi_{in}(0)|^{2+\delta} \le n^{-(2+\delta)} 2^{1+\delta}[EK_{1,h}^{2+\delta}(X) + (EK_{1,h}(X))^{2+\delta}].$$

Let $p_\delta = 1$, $\lambda_\delta = h/(2+\delta)$. We can show that

$$EK_{1,h}^{2+\delta}(X) = B_h(0)\int_0^\infty g(u, p_\delta, \lambda_\delta) f(u)du,$$

where $B_h(0) = 1/(2+\delta)h^{1+\delta}$. which, together with the result $\int_0^\infty g(u,p_\delta,\lambda_\delta)f(u)du = f(0)+o(1)$, implies $EK_{1,h}^{2+\delta}(X) = O(h^{-(1+\delta)})$. Hence

$$v_n^{-(2+\delta)}(0)\sum_{i=1}^n E\xi_{in}^{2+\delta}(0) = n^{-1-\delta}\cdot O\left((nh)^{(2+\delta)/2}\right)\cdot O(h^{-(1+\delta)}) = O((nh)^{-\delta/2}) = o(1)$$

by the assumption $nh \to \infty$. This completes the proof of (3.7).

Finally, let's consider the case of $x=0$, $f(0)=0$, $f'(0)\neq 0$. One can easily verify that

$$E\hat{f}_n(0) = f'(0)h + o(h), \quad \mathrm{Var}(\hat{f}_n(0)) = \frac{f'(0)}{4n} + o(n^{-1}).$$

For any $\delta > 0$, We can show that $EK_{1,h}^{2+\delta}(X) = O(h^{-\delta})$, Hence

$$v_n^{-(2+\delta)}(0)\sum_{i=1}^n E\xi_{in}^{2+\delta}(0) = n^{-1-\delta}\cdot O\left(n^{(2+\delta)/2}\right)\cdot O(h^{-\delta}) = O((nh^2)^{-\delta/2}) = o(1)$$

by the assumption $nh^2 \to \infty$. This completes the proof of (3.8). □

Proof of Theorem 3.3. Fix an $x>0$. To show the asymptotic normality of $\hat{m}_n(x)$, again we use the decomposition (5.11). We shall first show that $V_n(x)$ is asymptotically normal. For this purpose, let $\eta_{in} = n^{-1}K_{x/h+1,h}(X_i)\varepsilon_i$ 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 $E\eta_{in}^2 = [f(x)\sigma^2(x)/(2n^2\sqrt{x\pi h})[1 + o(1)]$. Therefore,

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

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 $\delta > 0$,

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

Hence

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

Hence, by the Lindeberg-Feller CLT, $s_n^{-1}V_n(x) \to_d N(0,1)$. Also, from the asymptotic results on $\hat{f}_n(x)$, $B_n(x)$, and $V_n(x)$, 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 $\sqrt{n\sqrt{h}} \cdot O_p(h^{1/4}/\sqrt{n}) = o_p(1)$, implies

$$f(x)s_n^{-1}[\hat{m}_n(x) - m(x) - hb(x) + o(h)] = s_n^{-1}V_n(x) \to_d N(0,1).$$

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

Proof of Theorem 3.4. Recall that $E\hat{f}_n(x) = \int_0^\infty g(u; p_1, \lambda_1)f(u)du$. Applying Lemma 5.1 with $k = 1$, $l = f$, and note that the boundedness of $f'(x)$ and $xf''(x)$ on $[a, b]$, we obtain

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

Hence $\sup_{a \le x \le b} |E\hat{f}_n(x) - f(x)| = O(h)$. So, we only need to show that $\hat{f}_n(x) - E\hat{f}_n(x) = o(\sqrt{\log n}/\sqrt{n\sqrt{h}})$. For this purpose, let $\xi_{in}(x) = n^{-1}[K_{x/h+1,h}(X_i) - EK_{x/h+1,h}(X_i)]$, 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 $k \ge 3$, $E|\xi_{1n}|^k \le c^{k-2}k!E\xi_{1n}^2$ for some c only depending on n.

Note that $K_{x/h+1,h}(X)$ can be written as

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

As a function of u, $u^{x/h}\exp(-u)$ attains its maximum at $u = x/h$. Therefore, by Stirling formula,

$$\begin{aligned} K_{x/h+1,h}(X) &\le \frac{1}{h\Gamma(x/h+1)} \left(\frac{x}{h}\right)^{x/h} \exp\left(-\frac{x}{h}\right) \\ &= \frac{(x/h)^{x/h}\exp(-x/h)}{h\sqrt{2\pi x/h}(x/h)^{x/h}\exp(-x/h)[1+o(1)]} \le \frac{c}{\sqrt{xh}} \end{aligned} \tag{5.20}$$

for some positive constant c. Therefore, for any $k \ge 2$, and α_n large enough,

$$\begin{aligned} E|\xi_{in}|^k &= n^{-k}E|K_{x/h+1,h}(X_i) - EK_{x/h+1,h}(X_i)|^k \\ &\le \left(\frac{c}{n\sqrt{xh}}\right)^{k-2} n^{-2}E|K_{x/h+1,h}(X_i) - EK_{x/h+1,h}(X_i)|^2 = \left(\frac{c}{n\sqrt{xh}}\right)^{k-2} E\xi_{in}^2. \end{aligned}$$

Let $v_n = \Big(\sum_{i=1}^n E\xi_{in}^2\Big)^{1/2}$, this immediately implies,

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

or

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

By (5.20), $v_n^2 = f(x)/2n\sqrt{xh}\sqrt{\pi} + o\Big(1/n\sqrt{h}\Big)$. This, together with the fact that $f(x)$ is bounded away from 0 and ∞ on $[a, b]$, implies

$$E\Big[\frac{\xi_{in}}{v_n}\Big]^k \leq k!\Big(\frac{c}{\sqrt{n\sqrt{h}}}\Big)^{k-2} E\Big[\frac{\xi_{in}}{v_n}\Big]^2.$$

Then, applying Bernstein inequality, for any positive number c,

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

Since $\log n/n\sqrt{h} \to 0$, 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

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

Therefore, by Borel-Cantelli Lemma, and $v_n^2 = O(1/n\sqrt{h})$, we have

$$\hat{f}_n(x) - E\hat{f}_n(x) = \sum_{i=1}^n \xi_{in} = O\left(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\right).$$

To bound the sum $\sum_{i=1}^n$ uniformly for $x \in [a, b]$, we partition the closed interval $[a, b]$ by the equally spaced points x_i, $i = 0, 1, \ldots, N_n$ such that $a = x_0 < x_1 < \cdots < x_{N_n} = b$ and $N_n = n^5$. It is easily seen that

$$P\Big(\max_{0\leq j\leq N_n}\Big|\sum_{i=1}^n \xi_{in}(x_j)\Big| > c\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\Big) \leq \frac{2N_n}{n^8} = \frac{2}{n^3}.$$

Borel-Cantelli Lemma thus implies that

$$\max_{0\leq j\leq N_n}\Big|\sum_{i=1}^n \xi_{in}(x_j)\Big| = O\Big(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\Big). \tag{5.21}$$

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

$$\xi_{in}(x) - \xi_{in}(x_j) = n^{-1}[K_{x/h+1,h}(X_i) - EK_{x/h+1,h}(X_i)] \\ -n^{-1}[K_{x/h+1,h}(X_i) - EK_{x/h+1,h}(X_i)].$$

Then a Taylor expansion of $K_{x/h+1,h}(X_i)$ at $x = x_j$ up to the first order leads to the following expression for the difference $K_{x/h+1,h}(X_i) - K_{x/h+1,h}(X_i)$:

$$(x - x_j)\left(\frac{X_i}{h}\right)^{\tilde{x}/h} \exp\left(-\frac{X_i}{h}\right)\left[-\frac{\Gamma'(\tilde{x}/h+1)}{\Gamma^2(\tilde{x}/h+1)h^2} + \frac{1}{\Gamma(\tilde{x}/h+1)h^2}\log\left(\frac{X_i}{h}\right)\right]$$

$$= \frac{x - x_j}{\Gamma(\tilde{x}/h+1)h^2}\left(\frac{X_i}{h}\right)^{\tilde{x}/h} \exp\left(-\frac{X_i}{h}\right)\left[-\frac{\Gamma'(\tilde{x}/h+1)}{\Gamma(\tilde{x}/h+1)} + \log\left(\frac{X_i}{h}\right)\right] \tag{5.22}$$

where $\tilde{x}$ is some value between x and x_j. At this stage, we need the following expansion for the derivative of Gamma function:

$$\Gamma'(x) = \Gamma(x)\left[-\frac{1}{x} - \gamma + x\sum_{k=1}^{\infty}\frac{1}{k(k+x)}\right], \tag{5.23}$$

where $\gamma \doteq 0.5772$ is the Euler constant. Apply (5.23) in (5.22), we have

$$\frac{x - x_j}{\Gamma(\tilde{x}/h+1)}\left(\frac{X_i}{h}\right)^{\tilde{x}/h} \exp\left(-\frac{X_i}{h}\right)\left[\frac{h}{\tilde{x}+h} + \gamma - \frac{\tilde{x}+h}{h}\sum_{k=1}^{\infty}\frac{1}{k(k+\tilde{x}/h+1)}\log\left(\frac{X_i}{h}\right)\right].$$

Using Stirling approximation, we have

$$\frac{1}{\Gamma(\tilde{x}/h+1)h^2}\left(\frac{X_i}{h}\right)^{\tilde{x}/h} \exp\left(-\frac{X_i}{h}\right) = \left(\frac{X_i}{\tilde{x}}\right)^{\tilde{x}/h} \exp\left[\frac{\tilde{x}}{h}\left(1 - \frac{X_i}{\tilde{x}}\right)\right]\frac{\sqrt{h}}{\sqrt{2\pi\tilde{x}}}[1 + o(1)].$$

Therefore, the right hand side of (5.22) can be written as the sum

$$(x - x_j)\left(\frac{X_i}{\tilde{x}}\right)^{\tilde{x}/h} \exp\left[\frac{\tilde{x}}{h}\left(1 - \frac{X_i}{\tilde{x}}\right)\right]\frac{\sqrt{h}}{h^2\sqrt{2\pi\tilde{x}}}[O(1) + O(h^{-1}) + O(\log h)]$$
$$+ (x - x_j)\left(\frac{X_i}{\tilde{x}}\right)^{\tilde{x}/h} \exp\left[\frac{\tilde{x}}{h}\left(1 - \frac{X_i}{\tilde{x}}\right)\right]\frac{\sqrt{h}}{h^2\sqrt{2\pi\tilde{x}}}\log\left(\frac{X_i}{\tilde{x}}\right)[1 + o(1)].$$

Note that $a < \tilde{x} < b$, so for h small enough, one can easily show that

$$\left|\left(\frac{X_i}{\tilde{x}}\right)^{\tilde{x}/h} \exp\left[\frac{\tilde{x}}{h}\left(1 - \frac{X_i}{\tilde{x}}\right)\right]\left[1 + \log\left(\frac{X_i}{\tilde{x}}\right)\right]\right| \le 2.$$

Therefore,

$$\left|K_{x/h+1,h}(X_i) - K_{x/h+1,h}(X_i)\right| \le \frac{c|x - x_j|}{h^2\sqrt{h}} \le \frac{c}{N_n h^2\sqrt{h}}$$

and $|\xi_{in}(x)-\xi_{in}(X_j)|\le 1/(nN_nh^2\sqrt{h})$. This, together with the choice $N_n=n^5$, implies that

$$\left|\sum_{i=1}^n \xi_{in}(x)-\sum_{i=1}^n \xi_{in}(X_j)\right| \le \frac{c}{N_nh^2\sqrt{h}} = o\left(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\right).$$

Finally, from (5.21) and (5.23), we obtain

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

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. To show the uniform convergence of $\hat{m}_n(x)$ over $[a,b]$, based on the decomposition (5.11), It suffice to show that

$$\sup_{x\in[a,b]}\Big|\frac{B_n(x)}{\hat{f}_n(x)}\Big| = O(h)+O\Big(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\Big), \tag{5.24}$$

$$\sup_{x\in[a,b]}\Big|\frac{V_n(x)}{\hat{f}_n(x)}\Big| = O(h)+O\Big(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\Big). \tag{5.25}$$

We shall prove (5.25) only. (5.24) can be proved similarly.

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,

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

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

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

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

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

Hence,

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

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

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

and for $k \ge 2$, $E(|\varepsilon_{i,n}^{d_n}|^k|X_i) \le 2^{k-2}d_n^{k-2}E(|\varepsilon_{i,n}^{d_n}|^2|X_i)$. Then from (5.20) and the boundedness of $\sigma^2(x)$ over $(0, \infty)$, we have

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

Because

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

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

$$P\Big(\Big|\sum_{i=1}^n K_{x/h+1,h}(X_i)\varepsilon_{i,2}^{d_n}\Big| \ge c\sqrt{\log n}\sqrt{\sum_{i=1}^n E\Big[K_{x/h+1,h}(X_i)\varepsilon_{i,2}^{d_n}\Big]^2}\Big) \le 2\exp(-c^2\log n/8).$$

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

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

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

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

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 3.3. In fact, the only major difference is that, instead of using (5.22), we should use the inequality

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

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

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

This concludes the proof of Theorem 3.5. □

Acknowledgement: The authors would like to thank the Editor and the referees for their critical comments that helped to substantially improve the presentation of the paper.

References

[1] Azzalini, A. and Bowman, A. W. (1990) A look at some data on the Old Faithful geyser. Applied Statistics, **39**, 357–365.

[2] Bagai,I. and Rao, B.L.S. (1996). Kernel type density estimates for positive valued random variables. Sankhya, **A57**, 56–67.

[3] Bouezmarni, T. and Rolin, J.M. (2003). Consistency of the beta kernel density function estimate. The Canadian Journal of Statistics, **31**(1), 89–98.

[4] Chaubey, Y., Laïb, N., Sen, A. (2010). Generalised kernel smoothing for non-negative stationary ergodic processes. Journal of Nonparametric Statistics, **22**, 973–997.

[5] Chaubey, Y., Li, J., Sen, A., and Sen, P.K. (2012). A new smooth density estimator for non-negative density etimator. J. Indian Statistical Association, **50**, 83–104.

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

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

[8] Chen, S.X. (2000b). Probability density function estimation using Gamma kernels. Ann. Inst. Statist. Math., **52**(3), 471–480.

[9] Chen, S.X. (2002). Local linear smoothers using asymmetric kernels. Ann. Inst. Statist. Math., **34**(1), 312–323.

[10] Fan, J. (1993), Local linear regression smoothers and their minimax efficiencies, The Annals of Statistics, **21**(1), 196–216.

[11] Jones, M.C. and Henderson, D.A. (2007). Kernel-type density estimation on the unit interval. Biometrika, **94**(4), 977–984.

[12] Marron, J.S. and Ruppert,D. (1994). Transformation to reduce boundary bias in kernel density estimate. J. Roy. Statist. Soc. Ser. B, **56**(4), 653–672.

[13] Mnatsakanov, R. and Sarkisian, K. (2012). Varying kernel density estimation on $\mathbb{R}_+$. Statistics & Probability Letters, **82**(7), 1337–1345.

[14] Parzen, E. (1962). On the estimation of a probability density function and the mode. Ann. Math. Statist. **33**, 1065–1076.

[15] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. **27**, 832–837.

[16] Ruppert, D. and Wand, M.P. (1992). Correcting for kurtosis in density estimation. Austral. J. Statist. **34**, 19–29.

[17] Scaillet. O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. Nonparametric Statistics, **16**(1-2), 217–226.

[18] Schucany, W.R. (1989). Locally optimal window widths for kernel density estimation with large samples. Statist. Probab. Lett., **7**, 401–405.

[19] Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformations in density estimation (with comments). JASA, **86**, 343–361.

[20] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Champan & Hall/CRC.