Technical Report No. 2/08, March 2008

A SMOOTH ESTIMATOR OF REGRESSION FUNCTION FOR NON-NEGATIVE DEPENDENT RANDOM VARIABLES

Yogendra P. Chaubey, Naâmane Laib and Arusharka Sen

A Smooth Estimator of Regression Function for Non-negative Dependent Random Variables

Yogendra P. CHAUBEY
Department of Mathematics and Statistics
Concordia University, Montreal, Canada H4B 1R6
E-mail:chaubey@alcor.concordia.ca

Naâmane LAIB[1]
L.S.T.A., University Paris 6, France
E-mail: nal@ccr.jussieu.fr

Arusharka SEN
Department of Mathematics and Statistics
Concordia University, Montreal, Canada H4B 1R6
E-mail:asen@mathstat.concordia.ca

January 17, 2008

Abstract. Commonly used kernel regression estimators may not provide admissible values of the regression function or its functionals at the boundaries, for regressions with restricted support. Any smoothing method will become less accurate near the boundary of the observation interval because fewer observations can be averaged, and thus variance or bias can be affected. Here, we adapt Chaubey *et al.* (2007)'s method of density estimation for nonnegative random variables to define a smooth estimator of the regression function. The estimator is based on a generalization of Hille's lemma and a perturbation idea. Its uniform consistency and asymptotic normality are obtained, for the sake of generality, under a stationary ergodic process assumption for the data . The asymptotic mean squared error is derived and the optimal value of smoothing parameter is also discussed. Graphical illustration of the proposed estimator are provided on simulated as well as real-life data.

AMS 1991 subject classifications: Primary 62G05, 62M20, secondary 60J15.

Key words and phrases: Ergodic processes, Hille's Lemma, gamma density function, martingale difference, normality, prediction, regression function.

---

1 Introduction

Various nonparametric estimators of regression function $m(\cdot)$ have been proposed in the literature, we may refer to Tran (1994) and Laïb (2005) and the references therein. Note however, that most of these methods may not provide admissible values of the regression, or its functionals at the boundaries for restricted support regressions. Near the boundary of the observation interval any smoothing method will become less accurate because fewer observations can be averaged and thus variance or bias can be affected. Although the usual kernel method may be used to estimate $m(\cdot)$, this method has two drawbacks. The first drawback concerns positive mass outside of support as shown by Silverman (1986) for the kernel density estimator, since this estimator can assign

[1]Corresponding author

positive mass to some $x \in (-\infty, 0)$. It can perform very well only for densities that are not far from Gaussian in shape (see, e.g., Wand, Marron and Ruppert (1991)). The second drawback of this estimator is its failure to consistently estimate discontinuity at the boundary, for regressions on $[0, +\infty)$ with $m(0) > 0$.

The boundary problem is of great importance, for instance, in econometrics where the range of the variable of interest in important models is not the whole real line. The boundary is usually at zero, and significant bias error occurs in the vicinity of zero. For instance, the income data for a country can have most of the density mass near zero because of high unemployment. Financial transaction data are typically highly dependent and often close or approximately equal to zero for frequently traded stocks. In the context of life testing and analysis, the associated random variables are typically nonnegative.

For i.i.d observations, several methods have been developed in the past to cope with the boundary error. See for instance Zhang *et al.* (1999), the reflection method of Hall and Wehrly (1991) and, in the setting of fixed-design regression, the generalized jackknifing technique of Rice (1984) [see also Härdle (1990), pages 130-132]. Boundary phenomena have also been studied by Gasser and Müller (1979) and Müller (1984).

In addition, there are a number of approaches to density estimation $f(\cdot)$ exclusively for non-negative data. For instance: the transformation method (e.g., Wand, Marron and Rupport (1991)); the Bagai and Prakasa Rao (1996) method which, unfortunately, uses only the first r order-statistics to estimate $f(x)$ if x lies between the r-th and $(r + 1)$-st order-statistics; the Chaubey and Sen (1996) method based on Hille's (1948) smoothing lemma; the Gamma-kernel estimator of Chen (1999) and the inverse-Gaussian kernel estimator of Scaillet (2004); the Chaubey *et al.* (2007) method based on a generalization of Hille's smoothing lemma, coupled with a perturbation idea to take care of the boundary bias.

Note also that most of the above papers deal with density or regression estimators in the setting of independent random variables. However, a great deal of data in econometrics, engineering and natural sciences, among other areas, occur in the form of time series in which observations are dependent.

In this paper we propose a smooth estimator of the regression function for nonnegative data. The estimator is obtained by adapting the Chaubey *et al.* (2007) method for density estimation based on generalized Hille's lemma and perturbation. Further, the data are assumed to be sampled from a stationary, ergodic process to allow maximum possible generality in the dependence structure. We avoid the widely used strong mixing condition and its variants as a dependence measure. For one thing, the calculation of probabilistic dependence measures is generally not easy because it involves the complicated manipulation of taking the supremum over two sigma algebras . Moreover, the mixing properties (strong or not strong) of a number of well known processes is still an open problem such as the AR(1)-GARCH(1, 1) process (see Lu and Linton (2005)). Additionally, many well-known processes are not strong mixing. For instance, Chernick (1981) and Andrews (1984) have given examples in which the first order linear autoregressive process with discrete valued random innovation is not strong mixing. In particular, if $(\epsilon_i)_{i\in\mathbb{Z}}$ is a sequence of independent Bernoulli random variables with parameter q, then the process $X_i = \rho X_{i-1} + \epsilon_i = \sum_{k=0}^{\infty} \rho^k \epsilon_{i-k}$, where $\rho \in (0; 1/2]$, is not strong mixing since $\alpha_n = 1/4$ for all n (see Andrews, 1984). The process (X_i) is an example of ergodic processes that do not fulfill the strong mixing property. In the same spirit, Guégan and Ladoucette (2001) show that some long memory processes with Gaussian innovation are ergodic without being strong mixing. Another example is given in Bosq (1998, pp 57-58) where the chaotic process of type $X_i = T(X_{i-1})$, with T a measurable real function, is shown to be ergodic but not strongly mixing.

In Section 2, we first derive a raw estimator without perturbation. It is then shown that this estimator, $m_n(x)$, can be *inconsistent* at $x = 0$ for $m(0)$ except in special cases. Following the idea of Chaubey *et al.* (2007), this motivates us to consider the perturbed version $\tilde{m}_n(x)$. Thus it appears that perturbation is indeed a very useful new idea to deal with boundary bias in the case of nonnegative data, which also avoids the complication of some of the rigorous boundary correction methods mentioned above. Section 3 is devoted to the study of asymptotic properties of the proposed estimator. We establish there the uniform almost sure convergence of the estimator $\tilde{m}_n(\cdot)$ when the observations are assumed to be only stationary and ergodic, so that the results hold for both mixing and non mixing processes. However, the asymptotic normality is established under a weaker dependence condition. In comparison to strong mixing this dependence condition appears sufficiently mild. Also, the asymptotic mean squared error is derived and the optimal choice of smoothing parameter is discussed. Section 4 deals with the generalization of our results to higher dimensional case. Section 5 is devoted to the application of our results to the construction of confidence bands for the functions $m(\cdot)$ as well as nonparametric predictors. In Section 6 we give some graphical illustration of the proposed estimator on simulated as well as real-life data, the latter pertaining to hardwood sapling height-growth in a boreal forest. The proofs are deferred to the Appendix. In this context, the martingale techniques play a vital role that allow us to obtain optimal results as in the i.i.d setting.

2 Smooth estimator of the regression function

Let $Z_i = (X_i, Y_i)_{i\in\mathbb{N}}$ be a $\mathbb{R}^+ \times \mathbb{R}^+$-valued strictly stationary ergodic sequence process defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Let ϕ be a Borelian function of $\mathbb{R}^+$ into $\mathbb{R}$ such that $\mathbb{E}(|\phi(Y_0)|) < \infty$. Let $m(x) = E(\phi(Y_0)|X_0 = x)$ be the conditional mean function of $\phi(Y_0)$ given $X_0 = x$ which is assumed to be bounded on $\mathbb{R}^+$.

The problem of interest is to construct a smooth estimator of the regression function $m(\cdot)$ based on data $Z_i, i = 1, \dots, n$. To this end, the following generalization of the Hille's Lemma will be used.

Lemma A (Lemma 1, Chapter VII.1, Feller 1965). *Let h be any bounded and continuous function. Let $g_{x,n}(\cdot)$, $n = 1, 2, \dots$ be a family of densities functions with mean $\mu_n(x)$ and variance $u_n^2(x)$ then we have as $\mu_n(x) \to x$ and $u_n(x) \to 0$*

$$\tilde{h}(x) = \int_{-\infty}^{\infty} h(t)g_{x,n}(t)dt \to h(x) \quad \text{as} \quad n \to \infty. \tag{2.1}$$

The convergence is uniform in every subinterval in which $u_n(x) \to 0$ and h is uniformly continuous.

Letting in (2.1), $h(t) = m(t)f(t)$ and suppose that $g_{x,n}(\cdot)$ be a density function satisfying $\int t g_{x,n}(t)dt = \mu_n(x) \to x$ and $\int (t - \mu_n(x))^2 g_{x,n}(t)dt = \sigma_n^2(x) \to 0$ as $n \to \infty$. This allows us to get

$$\int h(t)g_{x,n}(t)dt \to h(x) \quad \text{as } n \to \infty. \tag{2.2}$$

Observe that the left hand side of (2.2) can be written as $E_f\left(\phi(Y_0)g_{x,n}(X_0)\right)$, where the expectation is taken with respect to $f(\cdot)$, this motivated the introduction of the following estimator of $m(\cdot)$, that is

$$m_n(x) \;=\; \frac{n^{-1}\sum_{i=1}^n \phi(Y_i) g_{x,n}(X_i)}{n^{-1}\sum_{i=1}^n g_{x,n}(X_i)},$$

when the denominator is non equal 0. The function $g_{x,n}(\cdot)$ may be generated by considering a density function $q_v(x)$ on $[0,\infty)$ with mean 1 and variance v^2, giving $g_{(x,n)}(t) = \frac{1}{x} q_{v_n}(\frac{t}{x})$. The estimate of $m(x)$ is then given by

$$m_n(x) \;=\; \frac{n^{-1}\sum_{i=1}^n \phi(Y_i) Q_{x,v_n}(X_i)}{n^{-1}\sum_{i=1}^n Q_{x,v_n}(X_i)}, \tag{2.3}$$

where $Q_{x,v_n}(t) = \frac{1}{x} q_{v_n}(\frac{t}{x})$ is a density function on $[0,\infty)$ with x mean and variance $(xv_n)^2 \to 0$ as $n \to \infty$.

The above estimator, however, may not be defined at $x = 0$, except in cases where $m_n(0) = \lim_{x\to 0^+} m_n(x)$ exists. For instance, if $Q_{v_n,x}(\cdot)$ is a gamma density function with mean x and variance $(xv)^2_n$, defined for $x > 0$, by

$$Q_{x,v_n}(t) \;=\; \frac{1}{\beta_x^{\alpha_n}\,\Gamma(\alpha_n)} t^{\alpha_n - 1}\, e^{-\alpha_n t/x}, \text{ where } \alpha_n = 1/v_n^2,\ \beta_x = v_n^2 x. \tag{2.4}$$

Then, the limit $m_n(0)$ may be computed as follows

$$\begin{aligned} m_n(0) &= \lim_{x\to 0^+} \frac{\sum_{i=1}^n \phi(Y_{[i]}) X_{(i)}^{\alpha_n - 1}\ e^{-\alpha_n X_{(i)}/x}}{\sum_{i=1}^n X_{(i)}^{\alpha_n-1}\ e^{-\alpha_n X_{(i)}/x}} \\ &= \lim_{x\to 0^+} \frac{\sum_{i=1}^n \phi(Y_{[i]}) X_{(i)}^{\alpha_n - 1}\ e^{-\alpha_n [X_{(i)} - X_{(1)}]/x}}{\sum_{i=1}^n X_{(i)}^{\alpha_n-1} e^{-\alpha_n [X_{(i)} - X_{(1)}]/x}} \\ &= \lim_{x\to 0^+} \frac{\phi(Y_{[1]}) X_{(1)}^{\alpha_n-1} + \sum_{i=2}^n \phi(Y_{[i]}) X_{(i)}^{\alpha_n - 1}\ e^{-\alpha_n [X_{(i)} - X_{(1)}]/x}}{X_{(1)}^{\alpha_n-1} + \sum_{i=2}^n X_{(i)}^{\alpha_n-1}\ e^{-\alpha_n [X_{(i)} - X_{(1)}]/x}} \\ &= \phi(Y_{[1]}), \end{aligned}$$

where $X_{(i)}$ stands for the order statistic of X_i and $Y_{[i]}$ the corresponding concommitant, i.e., $Y_{[i]} = Y_j$ if $X_{(i)} = X_j$. However, in this case $m_n(0)$ does not consistently estimate $m(0)$.

To see this, consider the following example. Let (X_i, Y_i) be a sequence of i.i.d. r.v. with joint density $f(x,y) = e^{-y}$ for $y \geq x \geq 0$. Thus $f(x) = e^{-x}$, $f(y|x) = e^{-(y-x)}$, $m(x) = \int_x^\infty y f(y|x) dy = x + 1$ and $G_x(y) = P(Y \leq y | X = x) = 1 - e^{-y+x}$. Since for all $t > 0$

$$P(Y_{[1]} \leq t) = \int_{-\infty}^{\infty} G_x(t) f_{(1)}(x) I(t \geq x) dx,$$

where $f_{(1)}(\cdot)$ stands for the density of $X_{(1)}$ and $I(\cdot)$ the indicator function, then we have, when $\phi(Y_{[1]}) = Y_{[1]}$, that

$$\begin{aligned} P\left(\sqrt{n}(Y_{[1]} - m(0)) \leq t\right) &= n\int_0^\infty G_x\left(\frac{t}{\sqrt{n}} + m(0)\right)(1 - F(x))^{n-1} f(x) dx \\ &= n\int_0^{1+tn^{-1/2}} \left(1 - e^{-1-tn^{-1/2}+x}\right) e^{-(n+1)x} dx \\ &\to 1 - e^{-1} \quad as \quad n \to \infty. \end{aligned}$$

In this case, $m_n(0)$ does not consistently estimate $m(0) = 1$. This would be the case in general, unless the conditional distribution of Y, given $X = 0$, is degenerate.

To alleviate this situation we consider the following perturbed version of the above regression estimator

$$\tilde{m}_n(x) := m_n(x + \epsilon_n) = \frac{n^{-1} \sum_{=1}^{n} \phi(Y_i) Q_{x+\epsilon_n, v_n}(X_i)}{n^{-1} \sum_{=1}^{n} Q_{x+\epsilon_n, v_n}(X_i)}, \quad x \geq 0, \tag{2.5}$$

where $Q_{x+\epsilon_n,v_n}(t) = \frac{t}{x+\epsilon_n} q_{v_n}(\frac{1}{x+\epsilon_n})$ and ϵ_n goes to 0 at an appropriate (sufficiently slow) rate as $n \to \infty$.

In this paper, we focus on the special case where $Q_{v_n,x+\epsilon_n}(\cdot)$ is a gamma density function with mean $x + \epsilon_n$ and variance $v_n^2(x + \epsilon_n)^2$. Namely, for $x \geq 0$,

$$Q_{x+\epsilon_n,v_n}(t) = \frac{1}{\beta_{x+\epsilon_n}^{\alpha_n} \Gamma(\alpha_n)} t^{\alpha_n - 1} e^{-\alpha_n t/(x+\epsilon_n)}, \text{ where } \alpha_n = 1/v_n^2, \ \beta_{x+\epsilon_n} = v_n^2(x + \epsilon_n). \tag{2.6}$$

Gamma density is naturally asymmetric to cope with discontinuity at $t = 0$.

2.1 Notations and hypotheses

In order to state our results we introduce some notations. Let $\mathcal{F}_i$ be the σ-field generated by $((X_1, Y_1), \ldots, (X_i, Y_i))$ and $\mathcal{G}_i$ that generated by $((X_1, Y_1), \ldots, (X_i, Y_i), X_{i+1})$. For $i \in \mathbb{N}$, let $f(\cdot|\mathcal{F}_{i-1})$ be the conditional density of X_i given $\mathcal{F}_{i-1}$ and $f(\cdot)$ be the common density of the X_i's. Let $\mathcal{C}_0(\mathbb{R})$ be the space of continuous functions going to zero at infinity and $\| \cdot \|$ be the sup norm. From now on, set $J = [a, b] \subset \mathbb{R}^+$ with $0 \leq a < b$.The notation $\xrightarrow{\mathcal{D}}$ stands for the convergence in distribution of random variables. For a random variable ξ write $\xi \in \mathcal{L}^p$ $(p > 0)$ if $\|\xi\|_p := (E|\xi|^p)^{1/p} < \infty$ and define the projection $\mathcal{P}_k$ by $\mathcal{P}_k \xi := E(\xi|\mathcal{F}_k) - E(\xi|\mathcal{F}_{k-1})$, $k \in \mathbb{N}$.

Our results are stated under some assumptions we gather hereafter for easy reference

(A0) $v_n \to 0$ and $\epsilon_n \to 0$ as $n \to \infty$.

(A1) For all $i \in \mathbb{N}$, $f(\cdot) \in \mathcal{C}_0(\mathbb{R})$ and $f(\cdot| \mathcal{F}_{i-1}) \in \mathcal{C}_0(\mathbb{R})$ almost surely (a.s.)

(A2) The sequence $\{n^{-1} \sum_{i=1}^{n} f(x| \mathcal{F}_{i-1})\}$ converges uniformly in x to $f(x)$ almost surely.

(A3) $\sup\{f(x) : x \in [a, b], a > 0\} > 0$.

(A4) The conditional mean of $\phi(Y_i)$ given $\mathcal{G}_{i-1}$ only depends on X_i, that is , for all $i \geq 1$, $\mathrm{E}\left(\phi(Y_i) \middle| \mathcal{G}_{i-1}\right) = m(X_i)$.

(A5) There exists some $\gamma > 2$ such that $\max_{1 \leq i \leq n} \mathrm{E}\left(|\phi(Y)|^\gamma | \mathcal{G}_{i-1}\right) < \infty$ a.s.

Remark 1.

Assumptions (A1) and (A2) is justified by the work of Györfi and Lugosi (1992) where the authors have pointed out that the ergodic condition alone is not sufficient to ensure the L^1 consistency of kernel or histogram density estimates. A complementary assumption is therefore needed like the existence and the absolutely continuous almost surely of the conditional distribution. Conditions (A3) is very common in nonparametric estimation. (A4) is satisfied, for instance, by letting $Y_i = X_{i+1}$ where $\{X_i\}$ is a Markov process. As pointed by Györfi *et al.* (1998), condition (A4) is necessary for establishing the consistency of partitioning estimate. (A5) is very weaker than those proposed elsewhere in the literature.

3 Main Results

3.1 Uniform strong consistency

Theorems 1 below deals with the uniform consistency of the estimator $\tilde{m}_n(\cdot)$.

Theorem 1 *Assuming (A0)-(A5) hold, then we have*

$$\sup_{x\in[a,b]} |\tilde{m}_n(x) - m(x)| = 0 \ a.s. \quad as \quad n \to \infty.$$

3.2 Asymptotic Normality

Theorem 2 below delas with asymptotic normality for $\tilde{m}_n(\cdot)$.

Theorem 2 . *Let $W_{2+\delta}(X_i) := E[\phi^{2+\delta}(Y_i) \mid \mathcal{G}_{i-1}]$ for some $\delta > 0$. Assuming conditions (A0)-(A4) hold and that*

$$nv_n \to \infty \quad as\ n \to \infty \quad and \quad \max_i \sup_t f(t|\mathcal{F}_{i-1}) < \infty, \tag{3.1}$$

the functions $m(\cdot)$, $f(\cdot)$ and $W_{2+\delta}(\cdot)$ have bounded derivatives up to order two.

(i) If $f(x) > 0$ at given $x \in \mathbb{R}_*^+$, then

$$\sqrt{nv_n}(\tilde{m}_n(x) - m(x) - \tilde{B}_n(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \quad \sigma^2(x)\right), \ where\ \sigma^2(x) = \frac{1}{2\sqrt{\pi}} \frac{W_2(x) - m^2(x)}{xf(x)},$$

and $\tilde{B}_n(\cdot)$, which is defined in (7.3), stands for the bias term of $\tilde{m}_n(\cdot)$.

(ii) Suppose that

$$\sup_y \sum_{i=1}^{\infty} \|\mathcal{P}_1 f(y|\mathcal{F}_i)\|_2 < \infty, \tag{3.2}$$

$n^{1/2}v_n^{5/2} \to 0$ and $n^{1/2}v_n^{1/2}\epsilon_n \to 0$ as $n \to \infty$, then

$$\sqrt{nv_n}(\tilde{m}_n(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \quad \sigma^2(x)\right).$$

(iii) If $x = 0$ and if $\epsilon_n v_n \to 0$, $nv_n\epsilon_n \to \infty$, $n^{1/2}v_n^{5/2}\epsilon_n^{1/2} \to 0$ and $n^{1/2}v_n^{1/2}\epsilon_n^{3/2} \to 0$ as $n \to \infty$, then

$$\sqrt{nv_n\epsilon_n}(\tilde{m}_n(0) - m(0)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \quad \sigma_0^2(0)\right)$$

where $\sigma_0^2(0) = \frac{1}{2\sqrt{\pi}} \frac{W_2(0)-m^2(0)}{f(0)}$.

Remark 2

The condition (3.2) replace in some what the strong mixing condition and allows us to give an estimate of the convergence rate of the bias term $\tilde{B}_n(\cdot)$. It holds for linear as well as many nonlinear processes, such as threshold autoregressive models, AR models with conditionally heteroscedastic errors (see, Wu (2003) and Wu and Shao (2004)).

Example 1.Nonlinear models.

a) Let $d \geq 1$ be a fixed integer and consider the nonlinear $AR(d)$ model

$$X_n = R_{\epsilon_n}(X_{n-1}, \ldots, X_{n-d}), \tag{3.3}$$

where R is a bivariate measurable function and $\{X_n\}$ is a stationary process. For different forms of R in (3.3) one can obtain threshold autoregressive models (TAR, Tong (1990)), AR models with conditionally heteroscedastic errors (ARCH, Engle (1982)) and exponential autoregressive models (EAR, Haggan and Ozaki (1981)) among others. By iterating R in (3.3) one can see that the process X_n defined in (3.3) may be written as $X_n = F(\ldots, \epsilon_{n-1}, \epsilon_n)$, where F is a measurable function. The process $\{X_n\}$ is a stationary and causal process and represents a huge class of time series models. In the case where $d = 1$, the process $\{X_n\}$ admits a unique stationary distribution if

$$E(\log L_\epsilon) < 0, \quad E(L_\epsilon^\alpha) + E\left(|x_0 - R\epsilon(x_0)|^\alpha\right) < \infty, \quad \text{where} \quad L_\epsilon = \sup_{x \neq y} \frac{|R_\epsilon(x) - R_\epsilon(y)|}{|x - y|} \tag{3.4}$$

holds for some $\alpha > 0$ and x_0 (see, Diaconis and Freedman, 1999).

Let $f(u|X_n)$ be the conditional density of X_{n+1} at u given X_n and assume that $\sup_{u \in \mathbb{R}} |f(u|X_0)| < \infty$ and there exists C and $\beta > 0$ such that for all z and z' in $\mathbb{R}$,

$$\sup_{u \in \mathbb{R}} |f(u|z) - f(u|z')| \leq C|z - z'|^\beta.$$

By the analogous proof as that of Theorem 3 in Wu (2003) we have $\sup_{u \in \mathbb{R}} \|\mathcal{P}_0 f(u|X_n)\|_2 = O(r^n)$ for some $r \in]0, 1[$ and therefore condition (3.2) holds.

b) Letting $\phi(Y) = Y$ and $Y_i = X_i$ where X_i is generated following an ARCH-model:

$$X_i = \theta X_{i-1} + \sqrt{a_0 + a_1 X_{i-1}^2}\, \epsilon_i \tag{3.5}$$

where $a_0 \geq 0$ and $0 \leq a_1 < 1$, the sequence ϵ_i is i.i.d and for any $i \geq 1$, ϵ_i is independent of X_{i-1}. By (3.4) a sufficient condition of the existence of stationarity distribution is $E\left(\log(|\theta| + |a_1 \epsilon|\right) < 0$ and $E(|\epsilon|^\alpha) < \infty$.

Let f_ϵ and f'_ϵ be the density function of ϵ and its derivative. The conditional density of $X_i = z$ given $X_{i-1} = x$ is $f(z|x) = \frac{1}{\sqrt{a_0 + ax}} f_\epsilon(\frac{z - \theta x}{\sqrt{a_0 + ax}})$. Using theorem 3 of Wu (2003) one can see that the condition (3.2) is satisfied whenever $\sup_{z \in \mathbb{R}}[|z f'_\epsilon(z) + f_\epsilon(z)] < \infty$ and $\sup_{z \in \mathbb{R}} |f(z|x)| < \infty$.

Example 2. **Linear models.**

Let $X_n = \sum_{i=0}^\infty a_i \epsilon_{n-i}$, where $\sum_{i=0}^\infty |a_i| < \infty$, $E(\epsilon_0) = 0$ and $E(\epsilon_0^2) < \infty$. The process X_n includes many useful special cases such that the causal ARMA models. By the analogous proof as that of Theorem 4 in Wu (2003), we can show that (3.2) holds whenever $\sup_x |f_\epsilon(x)| < \infty$ and $\sup_x |f'(x)_\epsilon| < \infty$.

3.3 Asymptotic mean squares error (AMSE) of the regression estimator

Here we consider only asymptotic mean squared error (AMSE) of $\tilde{m}_n(x)$ computed at one single positive point x. In this case we may let $\epsilon_n = 0$, as perturbation is not needed away from the boundary $x = 0$. In a future paper we shall consider asymptotic mean integrated squared error (AMISE) as well as data-driven choice of both the smoothing parameters (ϵ_n, v_n) via an empirical, *cross-validation* function derived from AMISE.

The AMSE may be deduced from Theorem 2 as follows:

$$\text{AMSE}(\tilde{m}_n(x)) \quad = \quad \tilde{B}_n^2(x) + \frac{1}{nv_n}\sigma^2(x) \quad \text{for } x > 0.$$

Using (A1) and (A2) the bias $\tilde{B}_n(x)$ defined in (7.3) can be written, for n sufficiently large, as

$$\tilde{B}_n(x) \quad = \quad \frac{\int_0^\infty (m(x) - m(t))Q_{x+\epsilon_n,v_n}(t)f(t)dt}{\int_0^\infty Q_{x+\epsilon_n,v_n}(t)f(t)dt}.$$

One get then, by a Taylor expansion of order 2 of the functions $t \mapsto h(t) = m(t)f(t)$ and $t \mapsto f(t)$ that

$$\tilde{B}_n(x) = \frac{-\epsilon_n f'(x)m(x) + \frac{1}{2}(x^2v_n^2 + 2x\epsilon_n v_n^2)(-2f'(x)m'(x) - f(x)m''(x)) + o(v_n^2 + \epsilon_n)}{f(x) + \epsilon_n f'(x) + \frac{1}{2}(x^2v_n^2 + 2x\epsilon_n v_n^2)f''(x) + O(v_n^2 + \epsilon_n)}. \quad (3.6)$$

Here we consider only the case where $x > 0$. In this case the bias term can be approximated when $\epsilon_n = 0$ and $v_n \to 0$ by

$$\begin{aligned} \tilde{B}_n := B_n(x) \quad &\approx \quad \frac{(-m''(x)f(x) - 2m'(x)f'(x))\,\frac{x^2v_n^2}{2}}{f(x) + \frac{x^2v_n^2}{2}f''(x)} \\ &\approx \quad \frac{(-m''(x)f(x) - 2m'(x)f'(x))\,\frac{x^2v_n^2}{2}}{f(x)} \quad \text{as} \quad v_n \to 0. \end{aligned} \quad (3.7)$$

Thus, we have for n sufficiently large, that

$$AMSE(m_n(x)) \quad \approx \quad a(x)x^4v_n^4 + b(x)\frac{1}{nv_n}, \quad (3.8)$$

where

$$a(x) = \left[\frac{m''(x)f(x) + 2m'(x)f'(x)}{2fx)}\right]^2 x^4 \quad \text{and} \quad b_n(x) = \frac{1}{2\sqrt{\pi}}\frac{W_2(x) - m^2(x)}{xf(x)}. \quad (3.9)$$

The above result means that the bias square, as a function of v_n, is increasing whereas the variance decreasing.

Minimizing now the quantity AMSE with respect to v_n, one get the AMSE optimal bandwidth $v_{opt} = v_0$:

$$v_0 = \left(\frac{b(x)}{4a(x)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (3.10)$$

The optimal rate of the AMSE is thus given by

$$\begin{aligned} AMSE_{opt} &= a(x)x^4v_0^4 + \frac{b(x)}{nv_0} \\ &= \left(1+\frac{x^4}{4}\right)\left(\frac{1}{16\pi^2}C_1C_2^4\right)^{1/5} n^{-4/5}, \end{aligned} \tag{3.11}$$

where

$$C_1 = \left[\frac{m''(x)f(x)+2m'(x)f'(x)}{fx)}\right]^2 \quad \text{and} \quad C_2 = \frac{W_2(x)-m^2(x)}{f(x)}. \tag{3.12}$$

4 Generalization to the d-dimensional case

We briefly discuss a generalization of our result to the d-dimensional case. For $d \geq 1$, denote by $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$ a d-dimensional vector random variable defined on $\mathbb{R}^{+d}$. Let $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^{+d}$ and $\epsilon_n = (\epsilon_{1n}, \ldots, \epsilon_{dn})$ such that for any $1 \leq i \leq d$, $\epsilon_{in} \to 0$. Then for any $\mathbf{t} \in \mathbb{R}^{+d}$, the density function defined in (2.6) takes the forme

$$Q_{\mathbf{x}+\epsilon_n,v}(\mathbf{t}) = \frac{1}{(\prod_{i=1}^d \beta_{x_i+\epsilon_{in}})^\alpha \ (\Gamma(\alpha))^d}\left(\prod_{i=1}^d t_i\right)^{\alpha-1} e^{-\alpha\sum_{i=1}^d \frac{t_i}{x_i+\epsilon_{in}}}, \tag{4.1}$$

where $\alpha := \alpha_n = 1/v^2$, $\quad \beta_{x_i+\epsilon_{in}} = v^2(x_i+\epsilon_{in})$ and $v := v_n$.

Let $\mathbf{Z}_i = (\mathbf{X}_i, Y_i)_{i\in\mathbb{N}}$ be a $\mathbb{R}^{+d}\times\mathbb{R}^+$-valued strictly stationary ergodic sequence. Let ϕ be a Borelian function of $\mathbb{R}^+$ into $\mathbb{R}$. We estimate then $m(\cdot)$ by

$$\tilde{m}_n(\mathbf{x}) = \frac{n^{-1}\sum_{i=1}^n \phi(Y_i)Q_{\mathbf{x}+\epsilon_n,v}(\mathbf{X}_i)}{n^{-1}\sum_{i=1}^n Q_{\mathbf{x}+\epsilon_n,v_n}(\mathbf{X}_i)}. \tag{4.2}$$

We consider the following σ-algebra: $\mathfrak{F}_i = \sigma(\mathbf{Z}_1, \ldots, \mathbf{Z}_i)$ and $\mathfrak{G}_i = \sigma(\mathbf{Z}_1, \ldots, \mathbf{Z}_i, \mathbf{X}_{i+1})$.

For $i \in \mathbb{N}$, let $f_{\mathbf{X}_i}(\cdot|\mathfrak{F}_i)$ be the conditional density of $\mathbf{X}_i$ given $\mathfrak{F}_{i-1}$ and $f(\cdot)$ be the marginal density of $\mathbf{X}_i$. One can then state and prove the following theorem.

Theorem 3 . *Assuming conditions (A1)-(A4). Moreover, suppose that the functions $m(\cdot)$, $f(\cdot)$ and $W_{2+\delta}(\cdot)$ have bounded partial derivatives up to order d and*

$$v_n^d \to 0, \quad nv_n^d \to \infty \quad \text{and} \quad \max_i \sup_{\mathbf{t}} f(\mathbf{t}|\mathfrak{F}_{i-1}) < \infty. \tag{4.3}$$

Then we have for $f(\mathbf{x}) > 0$ at given $\mathbf{x} \in \mathbb{R}_*^{+d}$ that

(i) $\sqrt{nv_n^d}(\tilde{m}_n(\mathbf{x}) - m(\mathbf{x}) - \tilde{B}_n(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma^2(\mathbf{x})\right)$ *where* $\sigma^2(\mathbf{x}) = \frac{1}{(2\sqrt{\pi})^d}\frac{W_2(\mathbf{x})-m^2(\mathbf{x})}{(\prod_{i=1}^d x_i)f(\mathbf{x})}$

ii) Suppose that (3.2) holds and $n^{1/2}v_n^{5d/2} \to 0$ and $n^{1/2}v_n^{1d/2}\epsilon_n^d \to 0$ as $n \to \infty$, then

$$\sqrt{nv_n^d}(\tilde{m}_n(\mathbf{x}) - m(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \sigma^2(\mathbf{x})\right)$$

iii) If $x = 0$ and if $\epsilon_n^d v_n^d \to 0$, $nv_n^d \epsilon_n^d \to \infty$, $n^{1/2} v_n^{5d/2} \epsilon_n^{d/2} \to 0$ and $n^{1/2} v_n^{d/2} \epsilon_n^{3d/2} \to 0$ as $n \to \infty$, then we have

$$\sqrt{nv_n^d \epsilon_n^d}(\tilde{m}_n(\mathbf{0}) - m(\mathbf{0})) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \quad \sigma_0^2(\mathbf{0})\right) \quad \textit{where} \quad \sigma_0^2(\mathbf{0}) = \frac{1}{(2\sqrt{\pi})^d} \frac{W_2(\mathbf{0}) - m^2(\mathbf{0})}{f(\mathbf{0})}.$$

5 Applications

5.1 Confidence bounds

Using Theorem 2, the asymptotic $100(1-\alpha)\%$ confidence band for the function $m(\cdot)$ is given by

$$m_n(x) \pm c_\alpha \left(\frac{\sigma_n(x)}{nv_n}\right)^{1/2}, \quad x > 0,$$

where c_α is the upper α quantile of the distribution of $\mathcal{N}(0,1)$ and $\sigma_n(\cdot)$ is an appropriate estimate of $\sigma(\cdot)$.

5.2 Prediction in Markov time series

Let $\{U_i;\ i \in \mathbb{N}\}$ be a real-valued strictly stationary process. The prediction aims at evaluating U_{N+1} given $U_1, \ldots, U_N$. To this end, set $\mathbf{X}_i = (U_1, \ldots, U_{i+d-1})$ and $Y_i = U_{i+d}$, $i = 1, 2, \ldots, n$, where $n = N - d + 1$, d is here appropriately defined. Whenever $(U_i)_{i \geq 1}$ is a Markov process of order d, a theoritical predictor of U_{N+1} is given by $U_{N+1}^* = m(X_n)$. The predictor estimator of U_{N+1} is then $\hat{U}_{N+1} = \tilde{m}_n(X_n)$, where $\tilde{m}_n(\cdot)$ is the estimate of $m(\cdot)$ given by (2.5).

The following Corollary based on Theorem 1 gives the asymptotic behavior of the empirical error of prediction.

Corollary 1 *Under hypotheses of Theorem 1, then we*

$$\left|\hat{U}_{N+1} - U_{N+1}^*\right| \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad N \to \infty.$$

Corollary below, which is a consequence of Theorem 2, deals with the normality asymptotic of the empirical error of prediction.

Corollary 2 *Under the assumptions of Theorem 3, we have when $x > 0$*

$$\frac{\sqrt{Nv_N}}{\sigma(x)} \left(\hat{U}_{N+1} - U_{N+1}^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

6 Illustrations

We illustrate our method with two sets of simulated data, one each from IID and autoregressive models, as well as a real-life dataset on hardwood sapling height-growth:

IID data. Here $X_1, \ldots, X_n$ are generated as iid Exponential with expectation 1, and we consider two models for $Y_1, \ldots, Y_n$: a) $Y_i = 0.5(6 - 4X_i + X_i^2)\varepsilon_i$; b) $Y_i = \sin(1/X_i)\varepsilon_i,\ 1 \le i \le n$. Here $\varepsilon_i,\ 1 \le i \le n$, are taken to be i.i.d Weibull $(1, 2)$, i.e., with density $g(\varepsilon) = 2\varepsilon \exp(-\varepsilon^2),\ \varepsilon \ge 0$.

Figure-1 and Figure-2 illustrate our estimator for Model-a and Model-b respectively, and also provides a comparison with the usual kernel estimator. In both the figures we take $n = 200, v = cn^{-1/5}$ for $c = 0.2, 0.5$ and the perturbation-parameter $\epsilon = 0$ on the plot on the *left*, $\epsilon = 0.5v^2$ on the *right*. The choice of ϵ is based on the relation $\epsilon = O(v^2)$ established in Chaubey *et al.* (2007) for density estimation. The kernel estimator is based on the standard Normal kernel, where the bandwidth is chosen to be $h = 0.5n^{-1/5}$.

Figure-1 (Model-a) shows that $\tilde{m}_n(\cdot)$ with a low $v = 0.2n^{-1/5}$ is affected by noisy observations, as is the standard Normal kernel estimator, even with a high bandwidth. However, $\tilde{m}_n$ with a high $v = 0.5n^{-1/5}$ adapts well to the shape of the true regression. Moreover, the right-hand plot in Figure-1 shows that the effect of the large outlier near zero is reduced as ϵ is changed from zero to $0.5v^2$. In Figure-2 (Model-b) all the estimators are comparable.

Autoregressive data. Here $X_1, \ldots, X_n$ are generated as:

$$X_i = 0.5X_{i-1} + (\sqrt{0.2 + 0.1X_{i-1}^2})\varepsilon_i,\ X_0 \text{ Exponential } (1),$$

where $n = 200$, $\varepsilon_i,\ 1 \le i \le n$, are i.i.d Weibull $(1, 3)$, i.e., with density $g(\varepsilon) = 3\varepsilon^2 \exp(-\varepsilon^3),\ \varepsilon \ge 0$. The two models for $Y_1, \ldots, Y_n$, as well as the choice of smoothing parameters and kernel function, are exactly the same as the i.i.d case above. The illustration/comparison is provided in Figure-3 and Figure-4 for Model-a and Model-b respectively. The choice of $v,\ \epsilon$ here are the same as in the IID case above.

Figure-3 shows that $\tilde{m}_n(\cdot)$ with a low $v = 0.2n^{-1/5}$ is affected by noisy observations, as in the IID case. However, the standard Normal kernel estimator and $\tilde{m}_n(\cdot)$ with $v = 0.5n^{-1/5}$ are comparable in this case. In Figure-4 $\tilde{m}_n(\cdot)$ with low as well as high v detect the shape of the true regression quite well, while the kernel estimator remains essentially flat over the entire range.

Hardwood sapling data. We apply our method to data on initial height (X) versus 5-year height-growth (Y) of naturally-occurring hardwood saplings in gap areas of the boreal forest around Lake Duparquet in north-western Quebec. Both the initial height (as of 1998) and the height-growth (over 1998–2003) were obtained from multi-temporal LIDAR (LIght Detection And Ranging) surveys. (Data courtesy: Prof. Benoit St-Onge and Ms. Udayalakshmi Vepakomma, University of Quebec at Montreal.) All measurements are in meters, and the sample consists of $n = 94$ saplings.

Figure-5(a) gives the scatter-plot and our estimator $\tilde{m}_n(\cdot)$ along with the Standard Normal kernel estimator for comparison. The bandwidth and perturbation-parameters (v, ϵ) for our estimator, as well as the bandwidth for the kernel estimator, were chosen by trial-and-error through visual inspection of the fitted lines and the residuals (Figure-5(b)). We would like to mention two points: firstly, the kernel estimator required a bandwidth $(2.8n^{-1/5})$ that is 7 times that of $v = 0.4n^{-1/5}$ of $\tilde{m}_n(\cdot)$ for a comparably smooth fit; this indicates robustness of $\tilde{m}_n(\cdot)$ vis-a-vis the kernel estimator. Secondly, $\tilde{m}_n(\cdot)$ captures quite clearly the stabilization (i.e., approaching a constant level) of growth as initial height — an indicator of age — increases, as is to be expected, whereas the kernel estimator shows a downward trend.

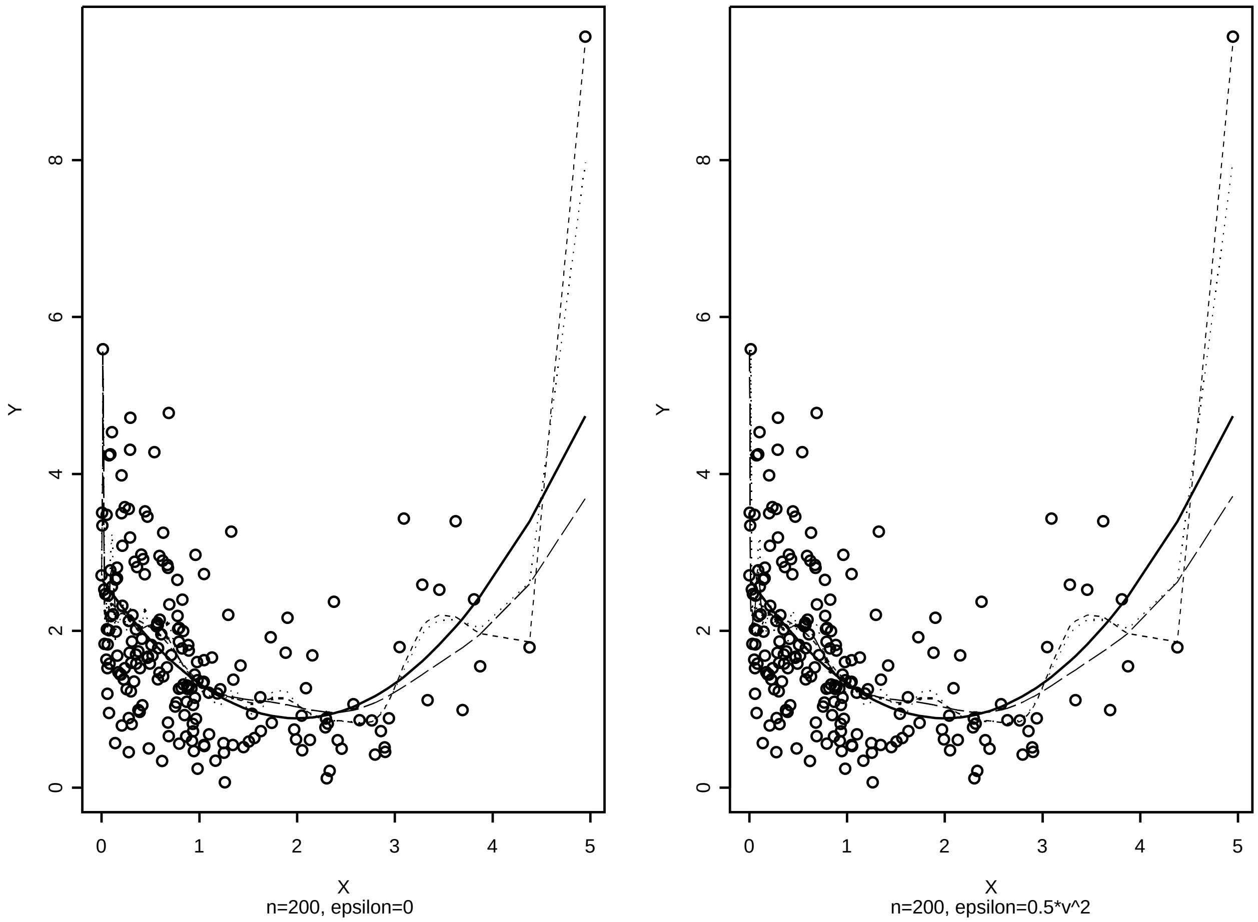

Figure 1: scatterplot and regression estimators for IID data with $Y = 0.5(6 - 4X + X^2)\varepsilon$: true regression (—), $\tilde{m}_n$ with $v = 0.2n^{-1/5}$ $(\cdots)$, $\tilde{m}_n$ with $v = 0.5n^{-1/5}$ (– –), standard Normal kernel with $h = 0.5n^{-1/5}$ (- - -)

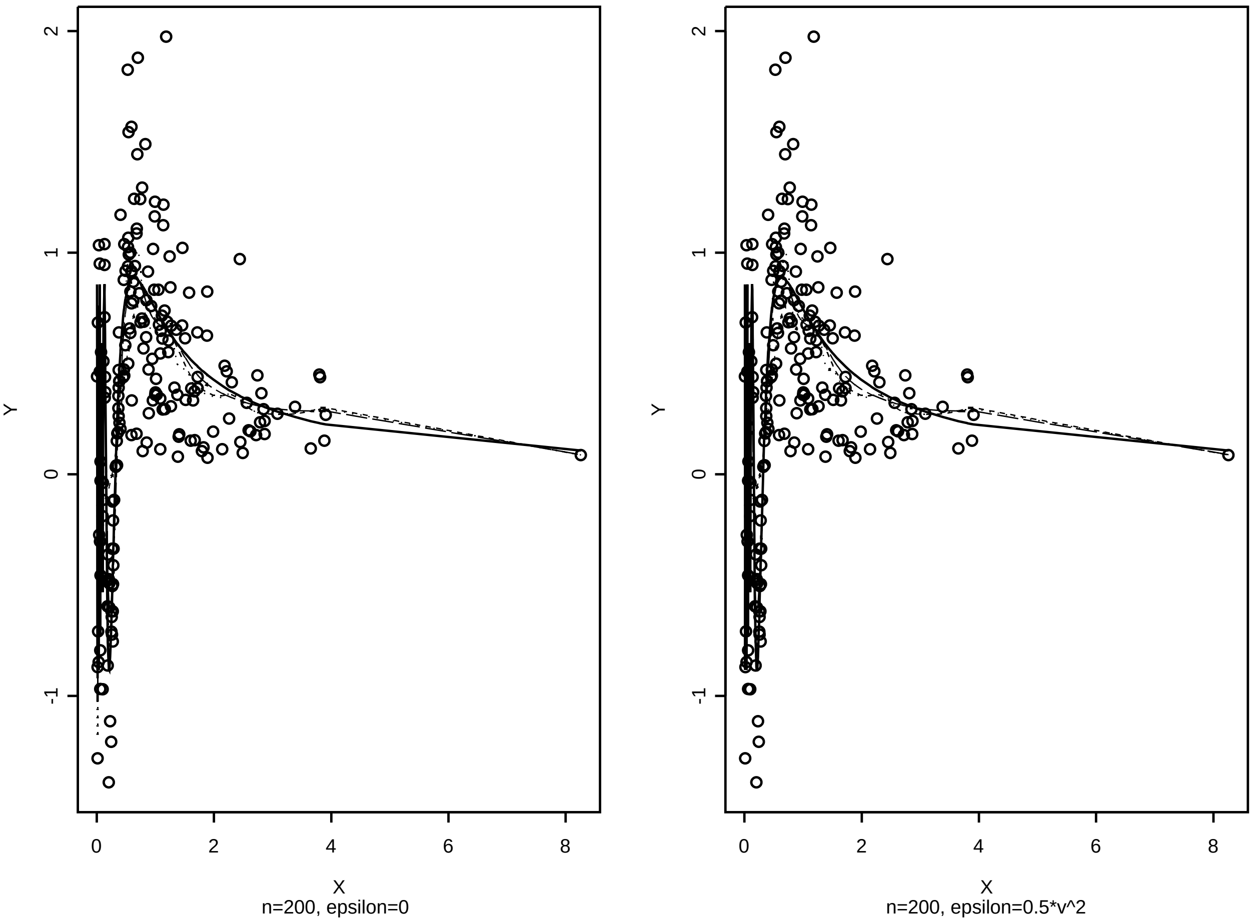

Figure 2: scatterplot and regression estimators for IID data with $Y = \sin(1/X)\varepsilon$: true regression (—), $\tilde{m}_n$ with $v = 0.2n^{-1/5}$ $(\cdots)$, $\tilde{m}_n$ with $v = 0.5n^{-1/5}$ (– –), standard Normal kernel with $h = 0.5n^{-1/5}$ (- - -)

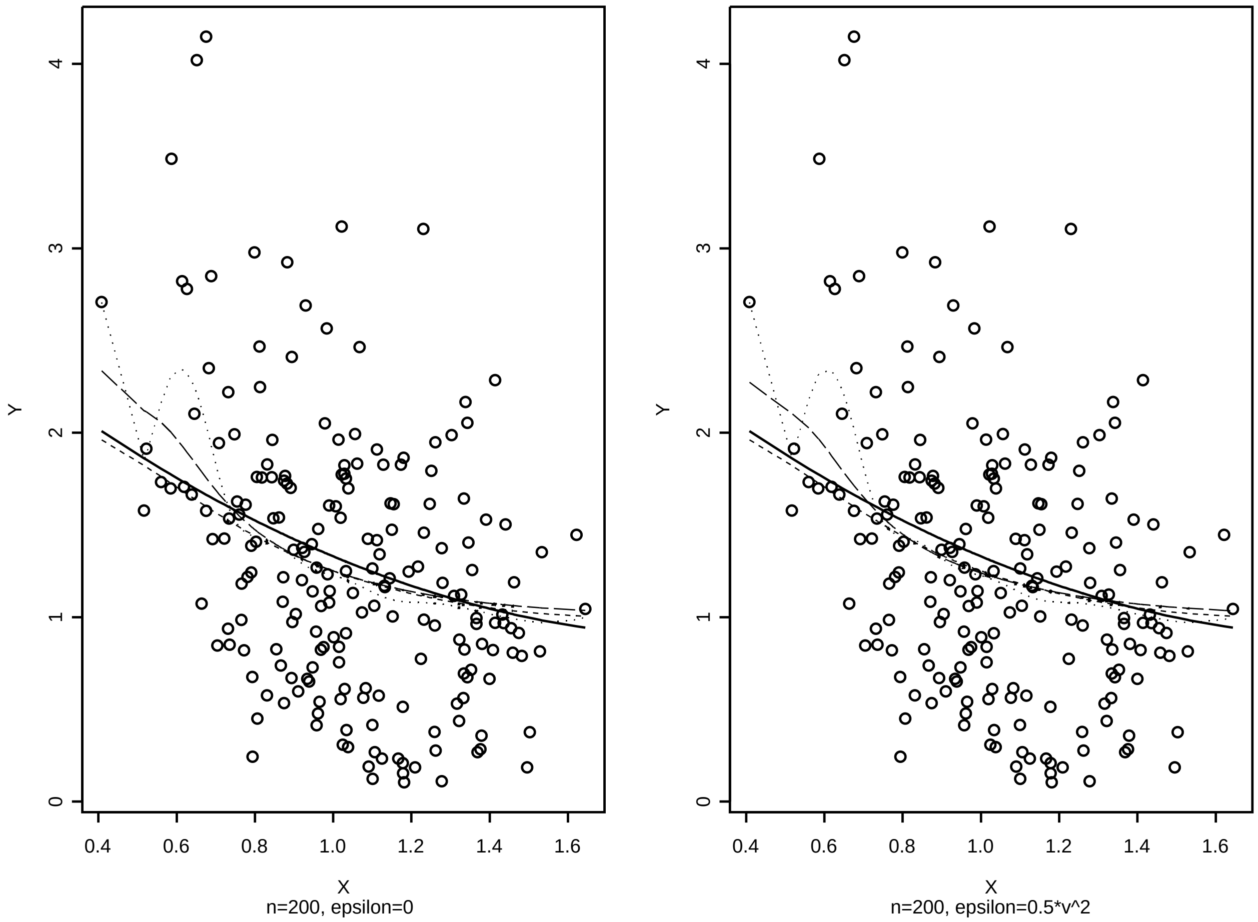

Figure 3: scatterplot and regression estimators for autoregressive data with $Y = 0.5(6-4X+X^2)\varepsilon$: true regression (—), $\tilde{m}_n$ with $v = 0.2n^{-1/5}$ $(\cdots)$, $\tilde{m}_n$ with $v = 0.5n^{-1/5}$ (– –), standard Normal kernel with $h = 0.5n^{-1/5}$ (- - -)

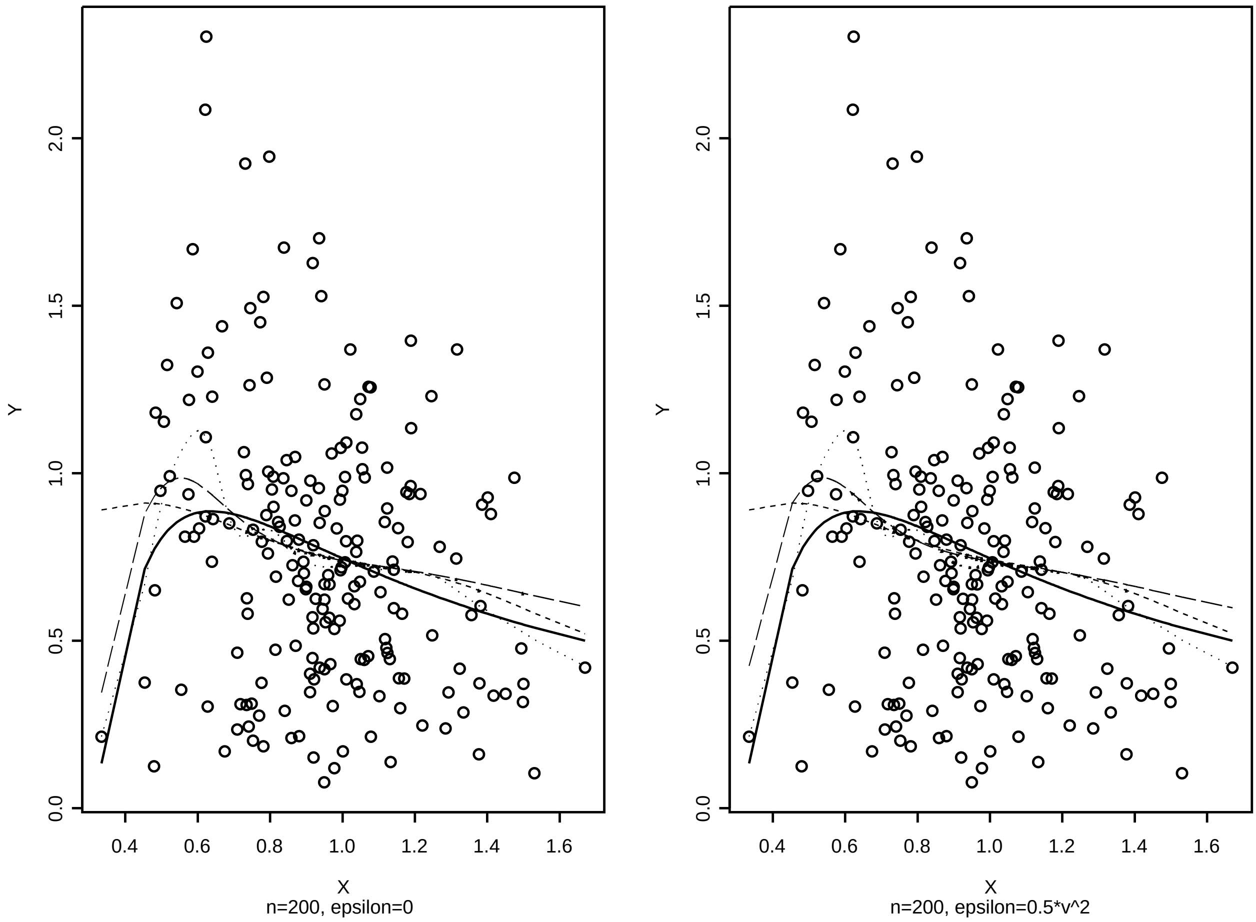

Figure 4: scatterplot and regression estimators for autoregressive data with $Y = \sin(1/X)\varepsilon$: true regression (—–), $\tilde{m}_n$ with $v = 0.2n^{-1/5}$ $(\cdots)$, $\tilde{m}_n$ with $v = 0.5n^{-1/5}$ (– –), standard Normal kernel with $h = 0.5n^{-1/5}$ (- - -)

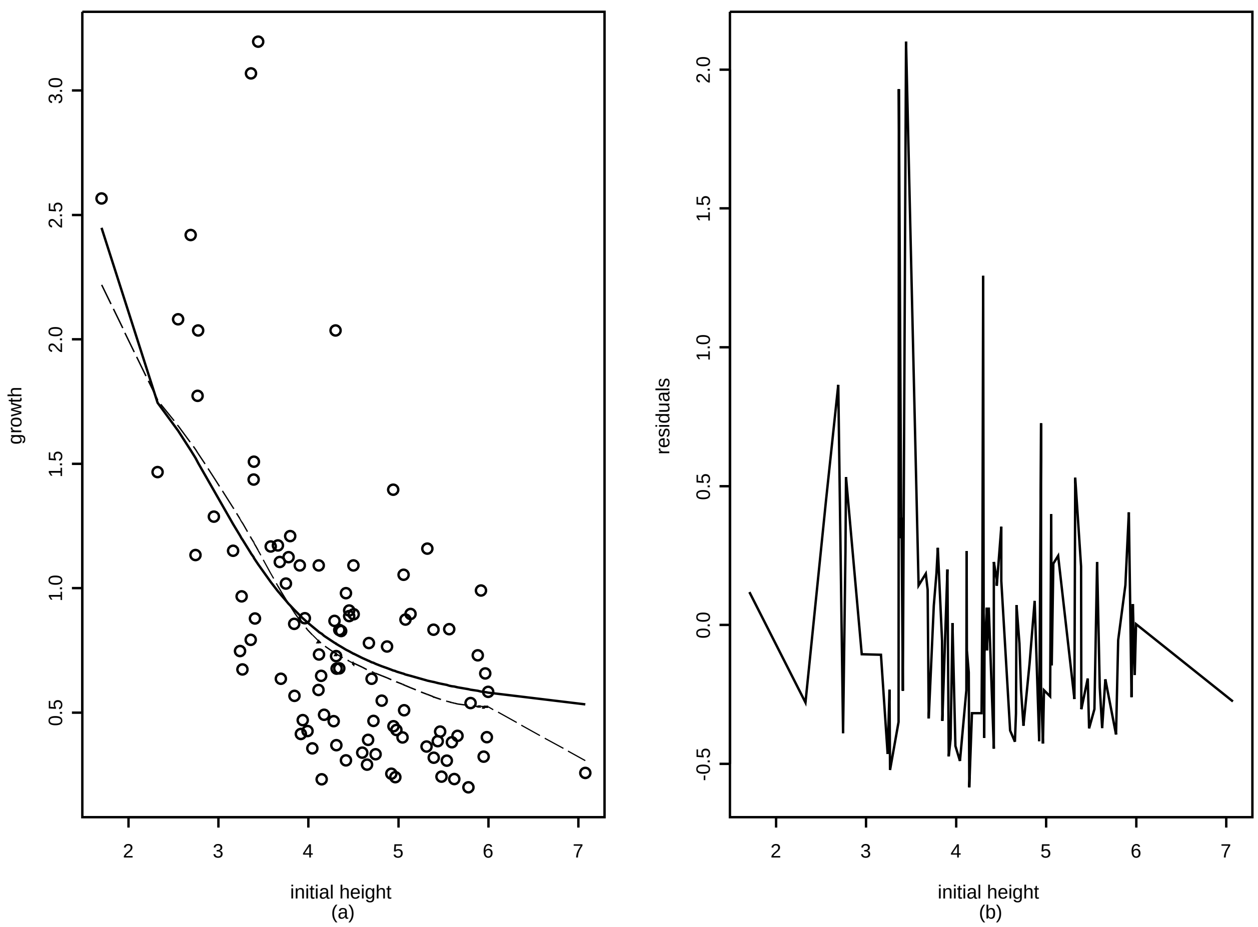

Figure 5: scatterplot and regression estimators for height-growth data: (a) $\tilde{m}_n$ with $v = 0.4n^{-1/5}$, $\epsilon = 0.5v^2$ (—), standard Normal kernel with $h = 2.8n^{-1/5}$ (– –); (b) line-plot of residuals corresponding to $\tilde{m}_n$

Acknowledgements: The authors gratefully acknowledge financial support from the Statistical Laboratory of CRM, Montreal (for N.Laib's trip to Montreal) and the NSERC Discovery grants (Y.Chaubey and A.Sen). The data-set used for illustration in Section 6 was kindly provided by B.St-Onge and U.Vepakomma of the UQAM.

7 Appendix: Proofs

This section gives detailed proofs. We start by two lemmas that we will be used in the sequel.

Lemma B (Laïb 1999). *Let* $\{(X_i, \mathcal{S}_i) : i \geq 1\}$ *be a sequence of martingale difference such that*

$|X_i| \leq B$ *a.s. for* $1 \leq i \leq n$. *For all* $\epsilon > 0$, *one has*

$$P\left\{\max_{1\leq i\leq n}\left|\sum_{j=1}^{i} X_j\right| > \epsilon\right\} \leq 2\exp\left(-\frac{\epsilon^2}{2nB^2}\right).$$

Lemma C (Wu, 2003). *For any* $\mathbf{y} \in \mathbb{R}^d$ *let* $H_n(\mathbf{y}) = \sum_{i=1}^n f(\mathbf{y}|\mathcal{F}_i) - nf(\mathbf{y})$. *Then condition (3.2) implies that* $\sup_y \|H_n(\mathbf{y})\|_2^2 = O(n)$.

In order to prove our results introduce some notations. For $x \in [a,b]$, let $x^+ = x + \epsilon_n$, $a_n = a + \epsilon_n$ and $b_n = b + \epsilon_n$. Let $h(x) = m(x)f(x)$ and $\tilde{m}_n(x) = m_n(x+\epsilon_n) := m_n(x^+)$. The estimator $\tilde{m}_n(x)$ of $m(x)$ can be written as

$$\begin{aligned}
\tilde{m}_n(x) &= \frac{h_n(x^+)}{f_n(x^+)}, \quad \text{where} \\
h_n(x^+) &= \frac{1}{n}\sum_{i=1}^n \phi(Y_i) Q_{x+\epsilon_n,v_n}(X_i) \quad \text{and} \quad f_n(x^+) = \frac{1}{n}\sum_{i=1}^n Q_{x+\epsilon_n,v_n}(X_i). \qquad (7.1)
\end{aligned}$$

Let

$$\overline{h}_n(x^+) = \frac{1}{n}\sum_{i=1}^n E\left[\phi(Y_i)Q_{x+\epsilon_n,v_n}(X_i)\right|\mathcal{F}_{i-1}] \quad \text{and} \quad \overline{f}_n(x^+) = \frac{1}{n}\sum_{i=1}^n E\left[Q_{x+\epsilon_n,v_n}(X_i)\right|\mathcal{F}_{i-1}]. \qquad (7.2)$$

We define the centralizing parameter

$$\tilde{B}_n(x) := \frac{\left[\overline{h}_n(x^+) - h(x)\right] - m(x)\left[\overline{f}_n(x^+) - f(x)\right]}{\overline{f}_n(x^+)} \qquad (7.3)$$

for the "bias" of $\tilde{m}_n(x)$. Then

$$\begin{aligned}
\tilde{m}_n(x) - m(x) - \tilde{B}_n(x) &= \frac{1}{f_n(x^+)}\left[(h_n(x^+) - \overline{h}_n(x^+))\right. \\
&\quad \left. -(m(x) + \tilde{B}_n(x))(f_n(x^+) - \overline{f}_n(x^+))\right], \qquad (7.4)
\end{aligned}$$

so that $\tilde{B}_n(\cdot)$ can be viewed as the "asymptotic bias" of $\tilde{m}_n(\cdot)$. The major thrust of the decomposition (7.4) is due to the fact that the summands of the term form a martingale difference.

We state and prove now the following results which give the uniform convergence of the bias term.

Proposition 1 *Assuming (A0)-(A4) hold, then we have*

$$\sup_{x\in[a,b]}\left|\tilde{B}_n(x)\right| = 0 \qquad a.s. \quad as \quad n \to +\infty.$$

Proof of Proposition 1. It suffices to show that $\overline{h}_n(x^+) - h(x)$ converges uniformly in x to 0 and $\overline{f}_n(x^+)$ is uniformly bounded over. Making use of (A4) and the law of iterated conditional expectation we can written

$$E\left[\phi(Y_i)Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] = E\left(E\left[\phi(Y_i)Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{G}_{i-1}\right]|\mathcal{F}_{i-1}\right) = E\left[Q_{x+\epsilon_n,v_n}(X_i)m(X_i)|\mathcal{F}_{i-1}\right].$$

Thus,

$$|\overline{h}_n(x^+) - h(x)| \leq \left| \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}^+} Q_{x+\epsilon_n, v_n}(t) m(t) f(t|\mathcal{F}_{i-1}) dt - h(x) \right|$$

$$\leq \quad \left\| \frac{1}{n} \sum_{i=1}^{n} f(\cdot|\mathcal{F}_{i-1}) - f \right\| \int_{\mathbb{R}^+} Q_{x+\epsilon_n, v_n}(t) m(t) dt + \left| \int_{\mathbb{R}^+} Q_{x+\epsilon_n, v_n}(t) h(t) dt - h(x) \right|. \quad (7.5)$$

By the Hille's Lemma and (A0), the second integral goes to 0 uniformly in x. The first term is bounded above by

$$\left\| \frac{1}{n} \sum_{i=1}^{n} f(\cdot|\mathcal{F}_{i-1}) - f \right\| . \sup_{x \in \mathbb{R}^+} |m(x)|,$$

which goes to 0 as $n \to \infty$ in view of (A2) and the fact that $m(\cdot)$ is bounded. By the same arguments we can conclude by Hille's Lemma, (A0) and (A2) that $\overline{f}_n(x^+)$ converges uniformly in x to $f(x)$ which is bounded over uniformly in x in view of (A3). □

The following Proposition gives an asymptotic lower bound for $\inf_{x \in J} |f_n(x^+)|$.

Proposition 2 *Assuming (A0)-(A3) hold, then we have*

$$(i) \qquad \sup_{x \in J} |f_n(x^+) - f(x)| = 0 \quad a.s. \quad as \quad n \to \infty$$
$$(ii) \qquad \inf_{x \in J} f_n(x^+) > 0 \quad a.s \quad as \quad n \to \infty.$$

Proof of Proposition 2. For (i) we have

$$|f_n(x^+) - f(x)| \leq |f_n(x^+) - \overline{f}_n(x^+)| + |\overline{f}_n(x^+) - f(x)|.$$

Making use of the same argument to prove Proposition 1, we can easily seen that the second term in the right hand side of the above inequality tends to 0 as $n \to \infty$. The first term converges also uniformly in x to 0 by the same arguments that used to prove Proposition 3 below. For (ii), we have for any $x \in J$,

$$\inf_{x \in J} |f_n(x^+)| \geq \inf_{x \in J} f(x) - \sup_{x \in J} |f_n(x^+) - f(x)|.$$

Then (ii) follows from (i) and condition (A3). □

The main task now is to establish the uniform almost sure convergence for $\overline{h}_n(x^+) - h(x)$. Making use of the Stirling's formula we can easily seen that, for any fixed x, the function $t \mapsto Q_{x+\epsilon_n, v_n}(t)$ is bounded above by $\frac{1}{\sqrt{2\pi(x+\epsilon_n)v_n}}$ for every $t \geq 0$ whenever $v_n \to 0$. By contrast, the function $\phi(y)$ is not necessarily bounded, it can thus be handled by a suitable truncation. To this end, let $M_n = \left\{ n \ln n \, [\ln \ln n]^{1+\zeta} \right\}^{1/\gamma}$, where ζ is a positive constant and γ is as in (A5). Note that the series $\sum_n M_n$ is convergent. Let us now define the following processes

$$h_n^b(x^+) \quad = \quad \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i) I\{|\phi(Y_i)| \leq M_n\} Q_{x+\epsilon_n, v_n}(X_i), \quad \text{and} \qquad (7.6)$$

$$\overline{h_n^b}(x^+) = \frac{1}{n}\sum_{i=1}^{n} E\left[\phi(Y_i)I\{|\phi(Y_i)| \le M_n\}Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right], \tag{7.7}$$

where I stands for the indicator function. We have

$$h_n(x^+) - \overline{h}_n(x^+) = (h_n(x^+) - h_n^b(x^+)) + (h_n^b(x^+) - \overline{h_n^b}(x^+)) + (\overline{h_n^b}(x^+) - \overline{h}_n(x^+)). \tag{7.8}$$

The asymptotic behavior of the three terms on the right hand side of (7.8) is given in the following results.

Lemma 1 *Assuming (A5) holds, then, for each ω outside a null set D, there exists a positive integer $n_0(\omega)$ such that $h_n(x^+) = h_n^b(x^+)$ for $n \ge n_0(\omega)$ and all $x \in \mathbb{R}^{+d}$.*

Proof of Lemma 1. The proof uses the summability of $M_n^{-\gamma}$ and arguments similar to those used by Roussas (1990). □

We deal now with the asymptotic behavior of the third term in (7.8).

Lemma 2 *Assuming (A2) and (A5) hold, then we have*

$$\sup_{x\in\mathbb{R}^+} |\overline{h_n^b}(x^+) - \overline{h}_n(x^+)| = O\left(M_n^{1-\gamma}\right) \quad a.s. \quad as \quad n \to \infty. \tag{7.9}$$

Proof of Lemma 2. We have by (A5) and the properties of conditional expectation that

$$\begin{aligned} E\left[\phi(Y_i)Q_{x+\epsilon_n,v_n}(X_i)I\{Y_i > M_n|\}\right] &\le M_n^{1-\gamma}E\left[|\phi(Y_i)|^{\gamma}Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] \\ &= M_n^{1-\gamma}E\left[Q_{x+\epsilon_n,v_n}(X_i)E\left[|\phi(Y_i)|^{\gamma}|\mathcal{G}_{i-1}\right]|\mathcal{F}_{i-1}\right] \\ &\le M_n^{1-\gamma}\max_{1\le i\le n} E\left(|\phi(Y_i)|^{\gamma}\mathcal{G}_{i-1}\right)E\left[Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] \\ &\le CM_n^{1-\gamma}\int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)f(t|\mathcal{F}_{i-1})dt. \end{aligned} \tag{7.10}$$

Therefore,

$$\begin{aligned} |\overline{h_n^b}(x^+) - \overline{h}_n(x^+)| \le\ & CM_n^{1-\gamma}\left\{\|\frac{1}{n}\sum_{i=1}^{n} f(\cdot|\mathcal{F}_{i-1}) - f(\cdot)\|\int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)dt\right. \\ &\left.+\int_{\mathbb{R}^+} Q_{x+\epsilon_n,v_n}(t)f(t)dt\right\}. \end{aligned} \tag{7.11}$$

The first member of (7.11) goes uniformly in x to 0 in view of (A2). The second one converges also uniformly in x, by Hille's Lemma and (A0), to $f(x)$ which is bounded. These imply (7.9). □

We study now the convergence of the main middle term on the right side of (7.8).

Proposition 3 *Let ν_n be a sequence of real number such that*

$$\nu_n \to \infty \quad and \quad \left(\frac{a_n}{b_n^2}\right)^{v_n} M_n v_n^{-3}\nu_n^{-1} \to 0 \quad as \quad n \to \infty. \tag{7.12}$$

Assuming (A0) holds and for any $\lambda > 0$

$$\sum_{n\geq 1} \nu_n \exp\left(-a_n^2 \pi \lambda^2 n M_n^{-2} v_n^2\right) < \infty. \tag{7.13}$$

Then, we have

$$\sup_{x\in[a,b]} |h_n^b(x^+) - \overline{h_n^b}(x^+)| = 0 \ a.s. \quad as \quad n \to \infty. \tag{7.14}$$

Remark 2. The condition (7.12) is satisfied if we choose, for instance, $\nu_n = \left[\left(\frac{a_n}{b_n^2}\right)^{v_n} M_n v_n^{-3} \log n\right] + 1$ whereas (7.13) holds true by taking $\lambda = \lambda_n = \sqrt{Cn}\frac{M_n}{a_n\sqrt{\pi n}v_n}$, where C is a large positive constant.

Proof of proposition 3. Divide the interval $[a,b]$ into subintervals each of length $\delta_n = (b-a)/\nu_n$. Since the set $J_n = \{x; |x| \leq |b-a|\}$ is compact, it can be covered by a finite number of bounded intervals with centers x_{nj} whose sides are of length δ_n. That is $J = [a,b] = \bigcup_{j=1}^{\nu_n} J_{nj}$, where

$$J_{nj} = \left\{x \ ; \ |x - x_{nj}| \leq (b-a)\nu_n^{-1}\right\}, \ j = 1, \ldots, \nu_n. \tag{7.15}$$

Let $V_n(x^+) = h_n^b(x^+) - \overline{h_n^b}(x^+)$, then we have, for $x_{nj} \in J_{nj}$, that

$$\begin{aligned}
\sup_{x\in J} |V_n(x^+)| &= \max_{1\leq j\leq \nu_n} \sup_{x\in J\cap J_{nj}} |V_n(x^+)| \\
&\leq \max_{1\leq j\leq \nu_n} \sup_{x\in J\cap J_{nj}} |V_n(x^+) - V_n(x_{nj}^+)| + \max_{1\leq j\leq \nu_n} |V_n(x_{nj}^+)| \\
&:= T_{1n} + T_{2n} + T_{3n},
\end{aligned}$$

where

$$T_{1n} = \max_{1\leq j\leq \nu_n} \sup_{x\in J_n\cap J_{nj}} |h_n^b(x^+) - h_n^b(x_{nj}^+)| \tag{7.16}$$

$$T_{2n} = \max_{1\leq j\leq \nu_n} \sup_{x\in J_n\cap J_{nj}} |\overline{h_n^b}(x^+) - \overline{h_n^b}(x_{nj}^+)| \tag{7.17}$$

$$T_{3n} = \max_{1\leq j\leq \nu_n} |h_n^b(x_{nj}^+) - \overline{h_n^b}(x_{nj}^+)|. \tag{7.18}$$

In order to give an upper bound of each term in the above inequalities we have to establish the following Lemmas.

Lemma 3 *Under (A0) we have*

$$\begin{aligned}
&(i) \qquad T_{1n} = O(\xi_n) \\
&(ii) \qquad T_{2n} = O(\xi_n) \quad with \quad \xi_n = C_1 a_n^{-4}\left(\frac{b_n^2}{a_n}\right)^{\alpha_n} \alpha_n^{3/2} M_n.\nu_n^{-1}
\end{aligned}$$

Proof of Lemma 3. We prove only (i), the proof of (ii) is similar. We have

$$|h_n^b(x^+) - h_n^b(x_{nj}^+)| \leq \frac{1}{n}\sum_{i=1}^{n} |\phi(Y_i)| I\{|\phi(Y_i)| \leq M_n\} \left|Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i)\right|.$$

Now observe that

$$Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i) = \frac{X_i^{\alpha_n-1}}{\Gamma(\alpha_n)}\left[\frac{e^{-\alpha_n X_i/x^+}}{\beta_{x^+}^{\alpha_n}} - \frac{e^{-\alpha_n X_i/x_{nj}^+}}{\beta_{x_{nj}^+}^{\alpha_n}}\right], \tag{7.19}$$

where $\beta_{x^+}^{\alpha} = (v_n^2 x^+)^{\alpha}$ and $\alpha_n = \frac{1}{v_n^2}$. The term in brackets in (7.19) can be written as

$$\frac{e^{-\alpha_n X_i/x^+} - e^{-\alpha_n X_i/x_{nj}^+}}{\beta_{x^+}^{\alpha_n}} + \frac{\left(\beta_{x_{nj}^+}^{\alpha_n} - \beta_{x^+}^{\alpha_n}\right) e^{-\alpha_n X_i/x_{nj}^+}}{\beta_{x^+}^{\alpha_n} \cdot \beta_{x_{nj}^+}^{\alpha_n}}. \tag{7.20}$$

Since for $c > 0$ and for any t, $(0 < a_0 \leq t \leq b)$, the function $f_c(t) = e^{-c/t}$ is a K_c lipshitz of order one with $K_c = \frac{c}{a_0^2} e^{-c/b}$, it follows, for all $x, x_n^j \in [a, b]$, that

$$|e^{-\alpha_n X_i/x^+} - e^{-\alpha_n X_i/x_{nj}^+}| \leq \frac{\alpha_n^2 X_i^2 e^{-\alpha_n X_i/b_n}}{a_n^2 x^+ \, x_{nj}^+} |x - x_{nj}|. \tag{7.21}$$

Moreover, making use of the mean value theorem, we can write, for x_* between x^+ and x_{nj}^+, that

$$\begin{aligned} |\beta_{x^+}^{\alpha_n} - \beta_{x_{nj}^+}^{\alpha_n}| &\leq \alpha_n v_n^{2\alpha_n} |x - x_{nj}| x_*^{\alpha_n - 1} \\ &\leq b_n^{\alpha_n-1} \alpha_n v_n^{2\alpha_n} |x - x_{nj}|. \end{aligned} \tag{7.22}$$

Combining (7.20), (7.21) and (7.22) we can then bound above the right hand of (7.19) by

$$\begin{aligned} & |Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i)| \\ \leq \; & \left[\frac{1}{a_n^4} \frac{\beta_{b_n}^{\alpha_n}}{\beta_{a_n}^{\alpha_n}} \alpha_n^2 X_i^{\alpha_n+1} Q_{b+\epsilon_n,v_n}(X_i) + \frac{\alpha_n v_n^{2\alpha_n} b_n{}^{\alpha_n-1}}{\beta_{a_n}^{\alpha_n}} X_i^{\alpha_n-1} Q_{a+\epsilon_n,v_n}(X_i)\right] |x_i - x_{nj}| \\ \leq \; & \left[\frac{1}{a_n^4}\left(\frac{b_n}{a_n}\right)^{\alpha_n} \alpha_n^2 X_i^{\alpha_n+1} Q_{b+\epsilon_n,v_n}(X_i) + \alpha_n b_n^{\alpha_n-1} a_n^{-2\alpha_n} X_i^{\alpha_n-1} Q_{a+\epsilon_n,v_n}(X_i)\right] |x_i - x_{nj}|. \end{aligned} \tag{7.23}$$

Making use of the Stirling's formula, we can see, for x fixed and $\tau \geq 0$, that the function $t \mapsto t^{\tau} Q_{x+\epsilon_n,v_n}(t)$ is bounded above by $\frac{(x+\epsilon_n)^{\tau-1}}{\sqrt{2\pi} v_n}$ whenever $v_n \to 0$. It follows that

$$|Q_{x+\epsilon_n,v_n}(X_i) - Q_{x_{nj}+\epsilon_n,v_n}(X_i)| \leq \frac{1 + a_n^4}{a_n^4 \sqrt{2\pi}} \left(\frac{b_n^2}{a_n}\right)^{\alpha_n} v_n^{-3} |x - x_{nj}|. \tag{7.24}$$

Hence,

$$T_{1n} \leq (b - a) \frac{1 + a_n^4}{a_n^4 \sqrt{2\pi}} \left(\frac{b_n^2}{a_n}\right)^{\alpha_n} v_n^{-3} M_n \nu_n^{-1} = O(\xi_n) \tag{7.25}$$

This completes the proof of Lemma 3. □

The following Lemma deals with the asymptotic behavior of T_{3n}.

Lemma 4 *Suppose that (A0) holds and that*

$$\sum_n \nu_n \exp\left(-a_n \pi \lambda^2 n v_n^2 M_n^{-2}\right) < \infty. \tag{7.26}$$

Then we have

$$T_{3n} = 0 \quad a.s \quad as \quad n \to \infty. \tag{7.27}$$

Proof of Lemma 4. The proof uses Lemma B. To this end write $|h_n^b(x_{nj}^+) - \overline{h_n^b}(x_{nj}^+)| = \sum_{i=1}^n L_n(x_{nj}^+)$, where $L_n(x_{nj}^+) = \frac{1}{n}\{\phi(Y_i) I\{|\phi(Y_i)| \leq M_n\} Q_{x_{nj}+\epsilon_n, v_n}(X_i)$. It is clear, for $x_{nj} \in [a,b]$, that

$$|L_n(x_{nj}^+)| \leq \frac{1}{\sqrt{2\pi x_{nj}^+ v_n}} n^{-1} M_n \leq \frac{1}{\sqrt{2\pi a_n v_n}} n^{-1} M_n,$$

whenever $v_n \to 0$. Moreover, for any fixed j, $1 \leq j \leq \nu_n$, $\left(L_n(x_{nj}^+), \mathcal{F}_i\right)$ is a bounded martingale difference, we can then apply Lemma B, to get for any $\lambda > 0$

$$P\{T_{3n} \geq \lambda\} \leq 2\nu_n \exp\left(-a_n \pi \lambda^2 n v_n^2 M_n^{-2}\right). \tag{7.28}$$

The result follows from Borel Cantelli's Lemma and condition (7.26). □

Proof of Theorem 1. The proof follows from decomposition (7.4), Propositions 1 to 3 and Lemmas 1 to 4. □.

Proof of Theorem 2.

(i) We have from (7.4) that for any $x > 0$

$$\sqrt{n v_n}\left(\tilde{m}_n(x) - m(x) - \tilde{B}_n(x)\right) = \frac{R_n(x^+)}{f_n(x^+)} - A_n(x^+), \tag{7.29}$$

where

$$\begin{aligned} R_n(x^+) &= \sqrt{n v_n}\left((h_n(x^+) - \overline{h_n}(x^+)) - m(x)(f_n(x^+) - \overline{f_n}(x^+))\right) \\ A_n(x^+) &= \sqrt{n v_n}\frac{\tilde{B}_n(x)(f_n(x^+) - \overline{f_n}(x^+))}{f_n(x^+)}. \end{aligned}$$

Let

$$\eta_{ni} = \left(\frac{v_n}{n}\right)^{1/2}\left[(\phi(Y_i) - m(x)) Q_{x+\epsilon_n, v_n}(X_i)\right] \quad \text{and} \quad \xi_{ni} = \eta_{ni} - E\left[\eta_{ni} | \mathcal{F}_{i-1}\right].$$

Then $R_n(x^+) = \sum_{i=1}^n \xi_{ni}$. Once the asymptotic normality of $R_n(x^+)$ is established, that of $\tilde{m}_n(x) - m(x) - \tilde{B}_n(x)$ follows from $A_n(x^+) \to 0$ in probability and $f_n(x^+) \to f(x)$ in probability as $n \to \infty$.

Lemma 5 below gives convergence in probability of $f_n(x^+)$ to $f(x)$.

Lemma 5 *Assuming (A0)-(A2) hold, then $f_n(x^+) \to f(x)$ in probability.*

Proof of Lemma 5. The result follows from a direct applications of Hille's Lemma combining with Lemma B. □

The following lemma gives the asymptotic behavior of $A_n(x^+)$.

Lemma 6 *Assuming (A0)-(A3) hold. If $f(x) > 0$ at a given $x \in \mathbb{R}_*^+$, then we have*

$$A_n(x^+) = o_P(1) \quad as \quad n \to \infty.$$

Proof of Lemma 6. Arguing as in the proof of Proposition 4 below by letting $m(x) = 0$ and $\phi(Y_i) \equiv 1$, we get for any $x > 0$, under condition (7.30), the following central limit theorem for the density estimator,

$$\sqrt{nv_n}(f_n(x^+) - \overline{f}_n(x^+) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \frac{f(x)}{2\sqrt{\pi x}}\right).$$

Thus, $f_n(x^+) - \overline{f}_n(x^+) = O_P(1/\sqrt{nv_n})$. It follows from Lemma 5 that $A_n(x^+) = O_P(1)|\tilde{B}_n(x)|$. We conclude by Proposition 1 that $A_n(x^+) = o_P(1)$. □

Proposition 4 *Let $W_{2+\delta}(X_i) = E[\phi^2(Y_i)|\mathcal{G}_{i-1}]$ be derivable at $X = x$ for some $\delta > 0$ and assume that $W_{2+\delta}(x)$ is bounded at a neighborhood of x. Moreover suppose that*

$$nv_n \to \infty \quad as\ n \to \infty \quad and \ \max_{1\leq i\leq n} \sup_t f(t|\mathcal{F}_{i-1}) < \infty. \tag{7.30}$$

Then we have for a given $x \in \mathbb{R}_+^*$ that

$$R_n(x^+) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ \tau^2(x)\right), \quad where \quad \tau^2(x) = \frac{f(x)}{2\sqrt{\pi x}}(W_2(x) - m^2(x)). \tag{7.31}$$

Proof of Proposition 4. Observe that for any fixed x, the summands in $R_n(x^+)$ form a triangular array stationary martingale differences with respect the sigma field $\mathcal{F}_{i-1}$, we can then apply a CLT for discrete-time arrays of real-valued martingales, as given for instance in Hall and Heyde (1980), to prove the asymptotic normality of $R_n(x^+)$. It suffices then to prove
$\sum_{i=1}^n E\left[\xi_{ni}^2|\mathcal{F}_{i-1}\right] \xrightarrow{\text{P}} \tau^2(x)$ and the Lindeberg condition
$nE\left[\xi_{ni}^2 I_{[|\xi_{ni}|>\varepsilon]}\right] = o(1)$ holds for any $\varepsilon > 0$.
In order to prove the first statement making use of condition (7.30) and the fact that $m(\cdot)$ is bounded, one get

$$\begin{aligned}
|E\left[\eta_{ni}|\mathcal{F}_{i-1}\right]| &= \left(\frac{v_n}{n}\right)^{1/2} E\left[(m(X_i) - m(x))Q_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] \\
&= \left(\frac{v_n}{n}\right)^{1/2} \int_{\mathbb{R}_*^+} (m(t) - m(x))Q_{x+\epsilon_n,v_n}(t)f(t|\mathcal{F}_{i-1})dt \\
&\leq C\left(\frac{v_n}{n}\right)^{1/2},
\end{aligned}$$

where $C = \max_i \sup_t m(t)f(t|\mathcal{F}_{i-1})$. Thus

$$\begin{aligned}
\left|\sum_{i=1}^n E\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] - \sum_{i=1}^n E\left[\xi_{ni}^2|\mathcal{F}_{i-1}\right]\right| &\leq \sum_{i=1}^n \left(E\left[\eta_{ni}|\mathcal{F}_{i-1}\right]\right)^2 \\
&\leq C^2.v_n \to 0 \quad \text{as} \quad n \to \infty. \qquad (7.32)
\end{aligned}$$

Consequently, we have only to prove

$$\sum_{i=1}^{n} E\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} \tau^2(x). \tag{7.33}$$

Observe now that

$$\begin{aligned}\sum_{i=1}^{n} E\left[\eta_{ni}^2|\mathcal{F}_{i-1}\right] &= \frac{v_n}{n}\sum_{i=1}^{n} E\left[W_2(X_i)Q^2_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] \\ &\quad -\frac{v_n}{n}\sum_{i=1}^{n} E\left[m(x)(2m(X_i)-m(x))Q^2_{x+\epsilon_n,v_n}(X_i)|\mathcal{F}_{i-1}\right] \\ &= J_{1n}+J_{2n}. \end{aligned} \tag{7.34}$$

The term J_{1n} can be be split as follows

$$J_{1n} = v_n \int_{\mathbb{R}_*^+} W_2(t)Q^2_{x+\epsilon_n,v_n}(t)\left[\frac{1}{n}\sum_{i=1}^{n} f(t|\mathcal{F}_{i-1}) - f(t)\right]dt + v_n\int_{\mathbb{R}_*^+} W_2(t)Q^2_{x+\epsilon_n,v_n}(t)f(t)dt. \tag{7.35}$$

By (A2) the term in brackets in (7.35) goes to 0 uniformly in t. Moreover, $v_n \int_{\mathbb{R}_*^+} W_2(t)Q^2_{x+\epsilon_n,v_n}(t)dt$ is bounded above by

$$v_n \sup_t Q_{x+\epsilon_n,v_n}(t)\int_{\mathbb{R}_*^+} W_2(t)Q_{x+\epsilon_n,v_n}(t)dt \approx \frac{1}{\sqrt{2\pi x}}W_2(x)$$

since by Hille's Lemma $\int_{\mathbb{R}_+^*} W_2(t)Q_{x+\epsilon_n,v_n}(t)dt \to W_2(x)$. This implies that the first member in J_{n1} goes to 0 as $n \to 0$. The second member in (7.35) can be split as

$$v_n\int_{\mathbb{R}_+^*}(W_2(t)f(t) - W_2(x)f(x)Q^2_{x+\epsilon_n,v_n}(t)dt + v_n\int_{\mathbb{R}_+^*} W_2(x)f(x)Q^2_{x+\epsilon_n,v_n}(t)dt. \tag{7.36}$$

Making use of a Taylor expansion of order one of the function $t \mapsto (W_2 f)(\cdot)$ and the fact that

$$\begin{aligned}\int_0^\infty t^p Q^m_{x+\epsilon_n,v_n}(t)dt &= \frac{(\frac{1}{v^2x^+})^{m/v^2}}{\left(\frac{m}{v^2x^+}\right)^{((m/v^2)+p+1-m)}}.\frac{\Gamma(m/v^2+p+1-m)}{\Gamma^m(1/v^2)} \\ &\approx \frac{1}{\sqrt{m(2\pi)^{m-1}}}\frac{1}{v^{m-1}(x+\epsilon_n)^{m-p-1}}\frac{1}{\sqrt{1-v^2(\frac{m-p-1}{m})}}, \quad \text{as } v\to 0, \end{aligned}\tag{7.37}$$

one can show that the first member in (7.36) tends to 0. Moreover, the second one is asymptotically equivalent, as $\epsilon_n \to 0$, to

$$J_{n1} \approx \frac{f(x)W_2(x)}{2\sqrt{\pi}\ x}. \tag{7.38}$$

We have now to study the asymptotic behavior of the second member in J_{n2}. Observe that

$$
\begin{aligned}
J_{n2} &= -v_n \int_{\mathbb{R}_*^+} (2m(t) - m(x)[n^{-1} \sum_{i=1}^{n} f(t|\mathcal{F}_{i-1}) - f(t)]dt \\
&- v_n m(x) \int_{\mathbb{R}_*^+} (m(t) - m(x)) Q^2_{x+\epsilon_n, v_n}(t) f(t) dt \\
&- v_n m(x) \int_{\mathbb{R}_*^+} (m(t)f(x) - m(x)f(x)) Q^2_{x+\epsilon_n, v_n}(t) f(t) dt \\
&- v_n m(x)^2 f(x) \int_{\mathbb{R}_*^+} Q^2_{x+\epsilon_n, v_n}(t) dt. \qquad (7.39)
\end{aligned}
$$

Using the same argument as above we can easily seen that J_{n2} is asymptotically equivalent, as $\epsilon_n \to 0$, to

$$
J_{n2} \approx -\frac{f(x)m^2(x)}{2\sqrt{\pi}\ x}. \qquad (7.40)
$$

Then (7.33) follows from (7.38) and (7.40).

The Lindeberg condition results from Corollary 9.5.2 in Chow and Teicher (1998) which implies that $nE[\xi^2_{ni} I(|\xi_{ni}| > \epsilon)] \le 4nE[\eta^2_{ni} I(|\eta_{ni}| > \varepsilon/2)]$.

Let $a > 1$ and $b > 1$ such that $\frac{1}{a} + \frac{1}{b} = 1$. Making use of Hölder and Markov inequalities one can write for all $\varepsilon > 0$

$$
E[\eta^2_{ni} I(|\eta_{ni}| > \varepsilon/2)] \le \frac{E|\eta_{ni}|^{2a}}{(\varepsilon/2)^{2a/b}}.
$$

Taking $2a = 2 + \delta$ we get

$$
\begin{aligned}
& 4nE[\eta^2_{ni} I(|\eta_{ni}| > \varepsilon/2)] = O(1).n^{-\delta} v_n^{(2+\delta)/2}.E[|(\phi(Y_i) - m(x))Q_{x+\epsilon_n, v_n}|^{2+\delta}] \\
\le\ & O(1).n^{-\delta/2} v_n^{(2+\delta)/2} \left[E(\phi(Y_i) Q_{x+\epsilon_n, v_n}(X_i))^{2+\delta} + E(m(x) Q_{x+\epsilon_n, v_n})^{2+\delta} \right] \\
=\ & O(1).n^{-\frac{\delta}{2}} v_n^{\frac{2+\delta}{2}} \left[\int_{\mathbb{R}^*_+} W_{2+\delta}(t) Q^{2+\delta}_{x+\epsilon_n, v_n}(t) f(t) dt + m^{2+\delta}(x) \int_{\mathbb{R}^*_+} Q^{2+\delta}_{x+\epsilon_n, v_n}(t) f(t) dt \right] (7.41)
\end{aligned}
$$

The first term in (7.41) can be written as

$$
O(1) n^{-\frac{\delta}{2}} v_n^{\frac{2+\delta}{2}} \left[\int_{\mathbb{R}^*_+} [W_{2+\delta}(t) f(t) - W_{2+\delta}(x) f(x)] Q^{2+\delta}_{x+\epsilon_n, v_n}(t) dt - \int_{\mathbb{R}^*_+} W_{2+\delta}(x) f(x) Q^{2+\delta}_{x+\epsilon_n, v_n}(t) \right] (7.42)
$$

Using the approximation formula given in (7.37), we get

$$
O(1) n^{-\delta/2} v_n^{(2+\delta)/2} \left[\int_{\mathbb{R}^*_+} W_{2+\delta}(x) f(x) Q^{2+\delta}_{x+\epsilon_n, v_n}(t) \right] = O(1)(nv_n)^{-\delta/2} \to 0 \quad \text{as} \quad n \to \infty. \quad (7.43)
$$

since $nv_n \to \infty$ as $n \to \infty$. By the continuity of the function $t \to W_{2+\delta}(t) f(t)$ one can show that the first member in (7.42) also goes to 0 as $n \to \infty$. Similarly one can show that the second term in (7.41) is asymptotically negligible. This completes the proof of part (i).

Part (ii). To prove (ii) we need to give an estimate of the convergence rate in probability of the bias term. This is the subject of the following lemma.

Lemma 7 *Suppose that (A0)-(A4) hold and the condition (3.2) is satisfied. Moreover, assuming that f and m have bounded derivatives up to order two. Then we have*

$$|\tilde{B}_n(x)| = O_{\mathbb{P}}\left(\max\left(\max(v_n^2, \epsilon_n), n^{-1}\right)\right) = O_{\mathbb{P}}\left(\max(v_n^2, \epsilon_n)\right). \tag{7.44}$$

Proof of Lemma 7. From (7.3), it suffices to give a convergence rate of $\overline{h}_n(x^+) - h(x)$. To this end write

$$\overline{h}_n(x^+) - h(x) = \left(\overline{h}_n(x^+) - Eh_n(x^+)\right) + \left(Eh_n(x^+) - h(x)\right). \tag{7.45}$$

Making use of (A4) one may write

$$\begin{aligned}\overline{h}_n(x^+) - Eh_n(x^+) &= \frac{1}{n}\int_{\mathbb{R}^+} m(t)Q_{x+\epsilon_n}(t)\left[\sum_{i=1}^{n} f(t \mid \mathcal{F}_{i-1}) - nf(t)\right]dt \\ &= \frac{1}{n}\int_{\mathbb{R}^+} m(t)Q_{x+\epsilon_n,\ v_n}(t)H_n(t)dt, \end{aligned} \tag{7.46}$$

where $H_n(t) = \sum_{i=1}^{n} f(t \mid \mathcal{F}_{i-1}) - nf(t)$. We have then by Cauchy inequality and Lemma C that

$$\begin{aligned} E\left[\left|\overline{h}_n(x^+) - Eh_n(x^+)\right|^2\right] &\leq \frac{1}{n^2}\cdot\left[E\left(H_n^2(t)\right)\right]\left(\int_{\mathbb{R}^+} m(t)Q_{x+\epsilon_n}(t)dt\right)^2 \\ &= O(n^{-1})\left(\int_{\mathbb{R}^+} m(t)Q_{x+\epsilon_n}(t)dt\right)^2. \end{aligned} \tag{7.47}$$

In order to deal with the second term in (7.47) recall that $Q_{x+\epsilon_n,v_n}(t) = \frac{1}{x+\epsilon_n} g_{\alpha,\beta}(\frac{t}{x+\epsilon_n})$, where $g_{\alpha,\beta}(\cdot)$ stands for the probability density function of the gamma distribution parameterized in terms of a shape parameter α and inverse scale parameter $\beta = \alpha = v_n^2$, which in turn, has mean equals 1 and variance v_n^2. Thus, we have, by a Taylor expansion one get

$$\begin{aligned} &\int_0^\infty Q_{x+\epsilon_n,v_n}(t)m(t)dt = \int_0^\infty g_{\alpha,\beta}(s)m((x+\epsilon_n)s)ds \\ = &\int_0^\infty g_{\alpha,\beta}(s)[m(x) + (x(s-1)+s\epsilon_n))m'(x) + \frac{(x(s-1)+s\epsilon_n))^2}{2}m''(x) \\ &+O\left((x(s-1)+s\epsilon_n))^2\right)]ds \\ = &\ O(1) + O(v_n^2) + O(\epsilon_n) + O\left(\max(v_n^2,\epsilon_n)\right) = O(1) + O\left(\max(v_n^2,\epsilon_n)\right) = O(1). \end{aligned} \tag{7.48}$$

Thus $E\left[\left|\overline{h}_n(x^+) - Eh_n(x^+)\right|^2\right] = O(n^{-1})$. By the same argument as above one can see that $(Eh_n(x^+) - h(x)) = O\left(\max(v_n^2, \epsilon_n)\right)$. These leads to the desired result.

Part (iii). The proof is similar of part (ii). □.

Proof of theorem 3. We only give the proof when $d = 2$. The proof of Lemma 5 and Lemma 6 still unchanging since the Hille's Lemma and Lemma A are also true on $\mathbb{R}^{+d}$. Let g now be a function defined on $\mathbb{R}^{+d}$ posses continuous bounded partial derivatives of order one at each point of an open set $S \subset \mathbb{R}^{+d}$. Then for each point (s,t), $(s,t) \neq (x+\epsilon_{n1}, y+\epsilon_{n2}) := (x^+, y^+)$, such that the line segment $L((s,t),(x^+,y^+))$ joining (s,t) and (x^+,y^+) in S, we have

$$\int_{\mathbb{R}^{+d}} g(s,t)Q^2_{(x+\epsilon_{n1},y+\epsilon_{n2}),v}(s,t)dsdt = I_{1n} + I_{2n}, \tag{7.49}$$

where

$$
\begin{aligned}
I_{1n} &= \int_{\mathbb{R}^{+d}} [g(s,t) - g(x^+, y^+)] Q^2_{(x+\epsilon_{n1}, y+\epsilon_{n2}),v}(s,t) ds dt \\
&\approx \int_{\mathbb{R}^{+d}} \left[(s - x^+) \frac{\partial g}{\partial x^+}(x^+, y^+) - (t - y^+) \frac{\partial g}{\partial y^+}(x^+, y^+) \right] Q^2_{x+\epsilon_{n1},v}(s) Q^2_{y+\epsilon_{n2},v}(t) ds dt \to 0,
\end{aligned}
$$

as $n \to \infty$, in view of the approximation formula (7.37) whenever the partial derivatives of g are bounded. Using again the approximation formula (7.37) and the continuity of g, we get

$$
I_{2n} = \int_{\mathbb{R}^{+d}} g(x^+, y^+) Q^2_{(x+\epsilon_{n1}, y+\epsilon_{n2}),v}(s,t) ds dt \approx g(x,y) \frac{1}{4\pi v^2 xy} \quad as \quad (\epsilon_{n1}, \epsilon_{n2}) \to (0,0).
$$

It suffices then to replace in the proof of proposition 4, v_n by v_n^d and to apply the above result with $g(s,t) = W_2(s,t) f(s,t)$ in (7.36) and $g(s,t) = m(s,t) f(s,t)$ in (7.39) and finally $g(s,t) = W_{2+\delta}(s,t) f(s,t)$ in (7.42). $\square$

References

Andrews, D.W.K., 1984. Non-strong mixing autoregressive processes. J. Appl. Probab. 21, 930-934.

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

Bosq, D., 1998. Nonparametric Statistics for Stochastic Processes. Lecture Note in Statistics. Springer, New York.

Bouezmarni, T. and Scaillet, O. (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* **21**, 390-412.

Chanda, K.C., 1974. Strong mixing of linear stochastic processes. J. Appl. Probab. 11, 401-408.

Chaubey, Y.P., Sen, A. and Sen, P.K. (2007). A new smooth density estimator for non-negative random variables. *Submitted.*

Chaubey, Y.P. and Sen, P.K. (1996). On smooth estimation of survival and density functions. *Statist. Decisions* **14**, 1-22.

Chen, S.X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* **31**, 131-145.

Chernick, M.R., 1981. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. Ann. Probab. 9, 145-149.

Chow, Y.S. and Teicher, H. (1998). *Probabilty Theory*, 2nd ed. Springer, New York.

Feller, W. (1965). *An Introduction to Probability Theory and its Applications*, Vol II. New York: John Wiley and Sons.

Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 41-76.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica.* **50**, 987-1007.

Gasser, T. and Müller, H.G. (1979). Kernel estimation of regression functions. In: *Smoothing Techniques for Curve Estimation*, eds. Gasser and Rosenblatt. Heidelberg: Springer-Verlag.

Guégan, D., Ladoucette, S., 2001. Non-mixing properties of long memory processes. C. R. Acad. Sci. Paris, 333, 373-376.

Györfi, L. and Lugosi, G. (1992). Kernel density estimation from ergodic sample is not universally consistent. *Comp. Stat. and Dta Analysis* **14**, 437-442.

Györfi, L., Morvai, G. and Yakowitz, S. (1998). Limits to consistent on line forecasting for ergodic time series. *IEEE Transactions on Information Theory* **44**, 868–892.

Hall, P. and Wehrly, T.E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric estimators. *J. Amer. Statistical Assn.*, **86**, 665-672.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press.

Härdle, W. (1991). *Smoothing Techniques with implementation in S.* Springer Verlag.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and its Application.* Academic Press. New York.

Hille, E. (1948). Functional Analysis and Semigroups. *Am. Math. Colloq. Pub 31*, New York.

Laïb, N., 2005. Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. J.S.P.I 134, 116-139.

Laïb, N. (1999). Exponential-type inequalities for martingale difference sequences: Application to nonparametric regression estimation. *Communication in Statistics, Theory and Methods*, *Series A*, **28**, 1565–1576.

Marron, J.S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc.*, Ser B, **56**, 653-671.

Müller, H.G. (1984). Boundary effects in nonparametric curve estimation models. In: *COMPSTAT, 84-89, Physica Verlag.*

Rice, J.A. (1984). Boundary modification for kernel regression. *Communication in Statistics, Series A* **13**, 893-900.

Roussas, G.G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes. Appl* **36**, 107-116.

Ruppert, D. and Wand, M.P. (1992). Correcting for kurtosis in density estimation. *Australian Journal of Statistics* **34**, 19-29.

Scaillet, O. (2004). Density esimation using inverse Gaussian and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics* **16**, 217-226.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* London: Chapman and Hall.

Tran, L.T. (1992). Kernel density estimation for linear processes. *Stochastic Process. Appl* **26**, 281-296.

Tran, L.T. (1994). Density estimation for time series by histograms. *Journal of Statistical Planing and inference* **40**, 61-79.

Tong, H., (1990). *Non-linear time series: a dynamical system approach.* Oxford University Press.

Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association* **86**, 343-361.

Wu, W.B. and Shao X. (2004). Limit theorems for iterated random functions. *J. Appl. Prob.* **41**, 425-436.

Wu, W.B, (2004). Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci.* USA 102, **40**, 14150–14154 (electronic). 62M10

Wu, W.B., (2003). Nonparametric estimation for stationary processes. *Technical Report*, No. **536**, University of Chicago.

Zhang., S., Karunamuni, R.J. and Jones, M.C. (1999). An improved estimator of the density function at the boundary. *J.Amer. Statistical Assn.* **94**, 1231-1241.

List of Recent Technical Reports

76. Michael A. Kouritzin and Wei Sun, *Rates for Branching Particle Approximations of Continuous–Discrete Filters*, December 2004

77. Rob Kaas and Qihe Tang, *Introducing a Dependence Structure to the Occurences in Studying Precise Large Deviations for the Total Claim Amount*, December 2004

78. Qihe Tang and Gurami Tsitsiashvili, *Finite and Infinite Time Ruin Probabilities in the Presence of Stochastic Returns on Investments*, December 2004

79. Alexander Melnikov and Victoria Skornyakova, *Efficient Hedging Methodology Applied to Equity–Linked Life Insurance*, February 2005

80. Qihe Tang, *The Finite Time Ruin Probability of the Compound Poisson Model with Constant Interest Force*, June 2005

81. Marc J. Goovaerts, Rob Kaas, Roger J.A. Laeven, Qihe Tang and Raluca Vernic, *The Tail Probability of Discounted Sums of Pareto–Like Losses in Insurance*, August 2005

82. Yogendra P. Chaubey and Haipeng Xu, *Smooth Estimation of Survival Functions under Mean Residual Life Ordering*, August 2005

83. Xiaowen Zhou, *Stepping–Stone Model with Circular Brownian Migration*, August 2005

84. José Garrido and Manuel Morales, *On the Expected Discounted Penalty Function for Lévy Risk Processes*, November 2005

85. Ze–Chun Hu, Zhi–Ming Ma and Wei Sun, *Extensions of Lévy–Khintchine Formula and Beurling–Deny Formula in Semi–Dirichlet Forms Setting*, February 2006

86. Ze–Chun Hu, Zhi–Ming Ma and Wei Sun, *Formulae of Beurling–Deny and Lejan for Non-Symmetric Dirichlet Forms*, February 2006

87. Ze–Chun Hu and Wei Sun, *A Note on Exponential Stability of the Non-Linear Filter for Denumerable Markov Chains*, February 2006

88. H. Brito–Santana, R. Rodríguez–Ramos, R. Guinovart–Díaz, J. Bravo–Castillero and F.J. Sabina, *Variational Bounds for Multiphase Transversely Isotropic Composites*, August 2006

89. José Garrido and Jun Zhou, *Credibility Theory for Generalized Linear and Mixed Models*, December 2006

90. Daniel Dufresne, José Garrido and Manuel Morales, *Fourier Inversion Formulas in Option Pricing and Insurance*, December 2006

91. Xiaowen Zhou, *A Superprocess Involving Both Branching and Coalescing*, December 2006

92. Yogendra P. Chaubey, Arusharka Sen and Pranab K. Sen, *A New Smooth Density Estimator for Non–Negative Random Variables*, January 2007

93. Md. Sharif Mozumder and José Garrido, *On the Relation between the Lévy Measure and the Jump Function of a Lévy Process*, October 2007

94. Arusharka Sen and Winfried Stute, *A Bi-Variate Kaplan-Meier Estimator via an Integral Equation*, October 2007

95. C. Sangüesa, *Uniform Error Bounds in Continuous Approximations of Nonnegative Random Variables Using Laplace Transforms*, January 2008

96. Yogendra P. Chaubey, Naâmane Laib and Arusharka Sen, *A Smooth Estimator of Regression Function for Non-negative Dependent Random Variables*, March 2008

Copies of technical reports can be requested from:

Dr. Wei Sun
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Blvd. West,
Montreal, QC, H3G 1M8 CANADA