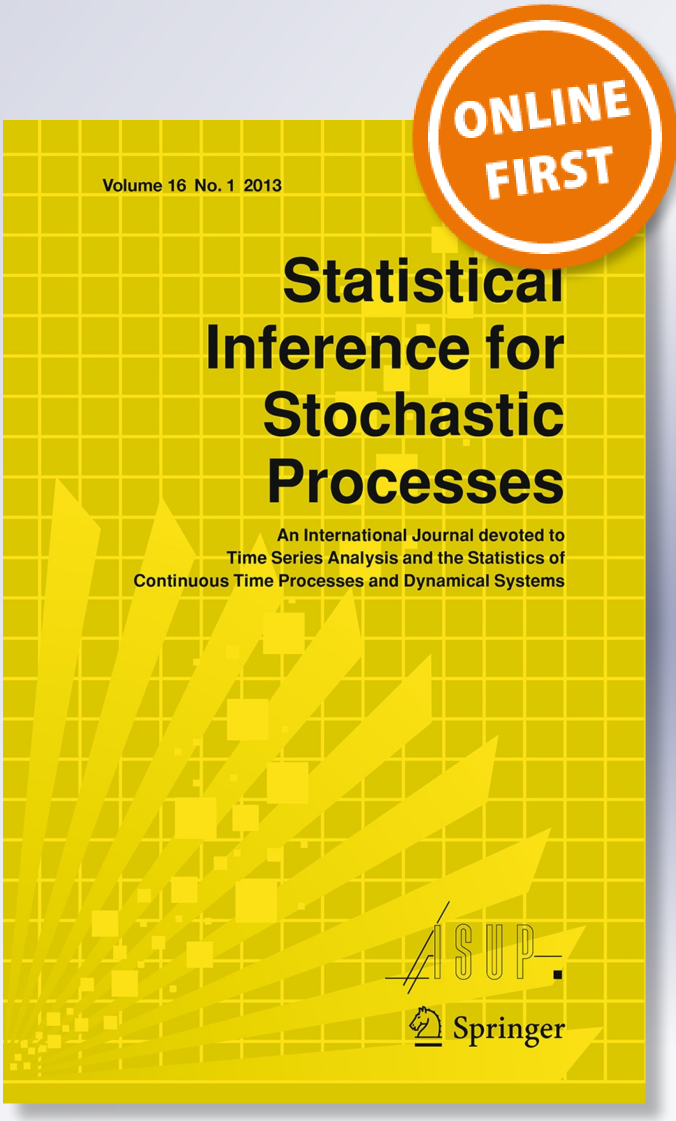
Asymptotic normality of recursive estimators under strong mixing conditions

Aboubacar Amiri

Statistical Inference for Stochastic Processes
An International Journal devoted to Time Series Analysis and the Statistics of Continuous Time Processes and Dynamical Systems

ISSN 1387-0874

Stat Inference Stoch Process
DOI 10.1007/s11203-013-9078-x



Springer

Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Springer

Author's personal copy

Stat Inference Stoch Process
DOI 10.1007/s11203-013-9078-x

Asymptotic normality of recursive estimators under strong mixing conditions

Aboubacar Amiri

Received: 12 June 2012 / Accepted: 25 January 2013
© Springer Science+Business Media Dordrecht 2013

Abstract The main purpose of this paper is to estimate the regression function by using a recursive nonparametric kernel approach. We derive the asymptotic normality for a general class of recursive kernel estimate of the regression function, under strong mixing conditions. Our purpose is to extend the work of Roussas and Tran (Ann Stat 20:98–120, 1992) concerning the Devroye–Wagner estimate.

Résumé Dans ce papier, nous nous intéressons à l'estimation de la fonction de régression par une approche non-paramétrique par noyau. Nous établissons la normalité asymptotique, pour une famille générale d'estimateurs récursifs à noyau de la fonction de régression, sous une hypothèse de forte mélangence. Notre rsultat généralise ainsi le résulttat de Roussas and Tran (Ann Stat 20:98–120, 1992) sur l'estimateur de Devroye–Wagner.

Keywords Recursive kernel estimators · Regression function · Strong mixing processes · Asymptotic normality

Mathematics Subject Classification 62G05 · 62G07 · 62G08

1 Introduction

In this paper we consider nonparametric sequential estimation of a regression functional, for dependent observations. Regression function estimation is an important issue in data analysis and remains a subject of hight interest, which covers many applied fields such as prediction, econometrics, decision theory, classification, communications and control systems. The literature on this topic is still growing and some relevant works on the subject include the monographs by Prakasa-Rao (1983), Gyorfi et al. (1989) and Yoshihara (1994), while more recent results are presented in, for example, the books by Györfi et al. (2002)

A. Amiri (✉)
Laboratoire EQUIPPE, Université Lille 3, Université Lille Nord de France,
EA 4018, Villeneuve d'Ascq, France
e-mail: aboubacar.amiri@univ-lille3.fr

Published online: 08 June 2013

Author's personal copy

and Bosq and Blanke (2007). Sequential estimation is achieved with the use of recursive estimators, typically kernel ones, and the purpose of this paper is to study a certain class of them. An estimator is said to be 'recursive' if its value calculated from the first n observations, say f_n, is only a function of f_{n-1} and the nth observation. In this way, the estimator can be updated with each new observation added to the database. This recursive property is clearly useful in sequential investigations and also for a fairly large sample size, since addition of a new observation means that the non-recursive estimators must be entirely recomputed. Besides, we are required to store extensive data in order to re-calculate them.

The first kernel recursive regression estimator was introduced by Ahmad and Lin (1976) taking the form

$$r_n^{AL}(x) := \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_i}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_i}\right)},$$

which is a recursive version of the Nadaraya-Watson estimate. Also, Devroye and Wagner (1980) propose the recursive estimator of the form

$$r_n^{DW}(x) := \frac{\sum_{i=1}^{n} \frac{Y_i}{h_i} K\left(\frac{x-X_i}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i} K\left(\frac{x-X_i}{h_i}\right)}.$$

In the literature $r_n^{AL}(x)$ and $r_n^{DW}(x)$ are respectively the so-called recursive and semi-recursive estimators. Various results on the latter estimators were established in an independent and identically distributed (i.i.d.) case, by many authors, we cite, among many others, Ahmad and Lin (1976), Devroye (1981), Greblecki and Pawlak (1987), Krzyżak (1992) and Walk (2001). In the dependent case, the majority of works are focused on Devroye-Wagner estimate. In a context of strong mixing processes, Roussas (1990) gave the uniform almost sure convergence for $r_n^{DW}(x)$, and Roussas and Tran (1992) showed its asymptotic normality. Under φ-mixing conditions, Qin (1995) have provided the asymptotic normality of $r_n^{DW}(x)$, and Wang and Liang (2004) have studied the almost uniform convergence for truncated versions of $r_n^{DW}(x)$ and $r_n^{AL}(x)$ in the same context. It should be noted that, unlike the iid case, more results are only obtained for $r_n^{DW}(x)$ in dependent case. In particular, no asymptotic normality has so far been established for $r_n^{AL}(x)$ in this context. Also we remark that, the approach used by Roussas and Tran (1992) to establish the asymptotic normality of $r_n^{DW}(x)$ cannot be generalized step by step to $r_n^{AL}(x)$. Indeed, the adaptation of their proof to $r_n^{AL}(x)$, needs to assume that the sequence $\frac{1}{n}\sum_{i=1}^{n}(h_i/h_n)^{2d}$ converges to a finite limit, for the study of a few covariance terms. The earlier condition is not satisfied by the popular choice $h_n = cn^{-\frac{1}{d+4}}$ for $d > 3$. Also, their proof uses the fact that for all $i = 1, \ldots, n$ $h_n < h_i$, while the same approach applied to $r_n^{AL}(x)$, leads to assume that $h_n > h_i$, which contradicts the optimal choice of h_n.

This paper deals with an extension of the work by Roussas and Tran (1992) to the general family of recursive estimators introduced by Amiri (2012), whose $r_n^{DW}(x)$ and $r_n^{AL}(x)$ are special cases. The paper is organized as follows. In the next section, we present our main assumptions and the results for regression estimation. The proof of the main result is postponed until Sect. 3.

Author's personal copy

2 Sequential regression estimation

2.1 Notation and assumptions

Let $\{(X_t, Y_t), t \in \mathbb{N}\}$ be a sequence of random variables on probability space $(\Omega, \mathcal{F}, P)$, taking values in $\mathbb{R}^d \times \mathbb{R}^{d'}$ $(d \geq 1, d' \geq 1)$, and having probability density function $f_{(X,Y)}$ with respect to the Lebesgue mesure. We assume that m is a Borelian function on $\mathbb{R}^{d'}$ into $\mathbb{R}$ such that $\omega \mapsto m^2(Y_t(\omega))$ is P-integrable, and define the regression function as

$$r(x) := \begin{cases} \mathrm{E}\left(m\left(Y_0\right)|X_0 = x\right) = \dfrac{\int_{\mathbb{R}^{d'}} m(y) f_{(X,Y)}(x,y)dy}{f(x)} := \frac{\varphi(x)}{f(x)}, \text{ if } f(x) > 0 \\ \mathrm{E}m(Y_0), \text{ if } f(x) = 0, \end{cases}$$

where f is the probability density function of X_0. Note that the transformation m is chosen by the statistician, leading to multiple choices of estimation. Typical examples of m are identity and polynomial functions to estimate respectively the usual regression and the conditional moments.

Throughout the paper we suppose that $f, \varphi \in C_d^2(b)$, where $C_d^2(b)$ denotes the set of twice-differentiable functions, with bounded second derivative. This condition is classical in the area of nonparametric estimation and has been used by Roussas and Tran (1992), Bosq and Blanke (2007), among others.

To estimate the functional $r(x)$, we consider the general family of kernel regression estimators introduced in Amiri (2012), defined by

$$r_n^{\ell}(x) := \frac{\sum_{i=1}^{n} \frac{m\left(Y_i\right)}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^{n} \frac{1}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right)}, \quad \ell \in [0, 1], \tag{1}$$

which can be computed recursively by

$$r_n^{\ell}(x) = \frac{\left(\sum_{i=1}^{n-1} h_i^{d(1-\ell)}\right) \varphi_{n-1}^{\ell}(x) + \left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right) m(Y_n) K_n^{\ell}\left(x - X_n\right)}{\left(\sum_{i=1}^{n-1} h_i^{d(1-\ell)}\right) f_{n-1}^{\ell}(x) + \left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right) K_n^{\ell}\left(x - X_n\right)},$$

where

$$\varphi_n^{\ell}(x) := \frac{1}{\sum_{i=1}^{n} h_i^{d(1-\ell)}} \sum_{i=1}^{n} \frac{m(Y_i)}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right), \quad f_n^{\ell}(x) := \frac{1}{\sum_{i=1}^{n} h_i^{d(1-\ell)}} \sum_{i=1}^{n} \frac{1}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right),$$

and $K_i^{\ell}(\cdot) := \frac{1}{h_i^{d\ell} \sum_{j=1}^{i} h_j^{d(1-\ell)}} K\left(\frac{\cdot}{h_i}\right)$. Our class of estimates includes the popular kernel recursive estimators $r_n^{AL}(x)$ and $r_n^{DW}(x)$, corresponding to the cases $\ell = 0$ and $\ell = 1$, respectively.

At this point, we can make some assumptions and provide the main theorem. Throughout this paper the kernel K is assumed to satisfy the following conditions.

Author's personal copy

Assumption H1 (i) $K : \mathbb{R}^d \mapsto \mathbb{R}$ is bounded, symmetric and positive function such that $\int_{\mathbb{R}^d} K(t)dt = 1$;

(ii) $\lim_{\|x\| \to +\infty} \|x\|^d K(x) = 0$;

(iii) $\int_{\mathbb{R}^d} |v_i v_j| K(v) dv < \infty, \;\; i, \;\; j = 1, \ldots, d.$

Assume the sequence h_n satisfies the following conditions.

Assumption H2 (i) $h_n \downarrow 0, nh_n^{d+2} \to \infty$;

(ii) For all $r \in (-\infty, d+2]$, $B_{n,r} := \frac{1}{n} \sum_{i=1}^{n} \left(\frac{h_i}{h_n} \right)^r \to \beta_r > 0$ as $n \to \infty$;

(iii) For each sequence of integers u_n and v_n such that $u_n \sim v_n$, then $h_{u_n} \sim h_{v_n}$. [1]

Assumption H3 (i) The process (X_t) is α-mixing with

$$\alpha_X(k) \leq \gamma k^{-\rho}, \;\; k \geq 1, \gamma > 0 \text{ and } \rho > \max\left(2, \frac{d+2}{2}\right);$$

(ii) For each couple $(s, t), \;\; s \neq t$, the random vector (X_s, X_t) admits a probability density function $f_{(X_s, X_t)}$ such that $\sup_{|s-t| \geq 1} \| \; g_{s,t} \; \|_\infty < \infty$, where $g_{s,t}(\cdot, \cdot) := f_{(X_s, X_t)}(\cdot, \cdot) - f(\cdot) f(\cdot)$.

Assumption H4 (i) The function $\mathrm{E}\left(m^2(Y)|X_0 = \cdot\right) f(\cdot)$ is both continuous and bounded away from zero at x;

(ii) There exist $\lambda > 0, \theta > 0$ such that $\mathrm{E}\exp(\lambda |m(Y_0)|^\theta) < \infty$;

(iii) For each $k \neq k'$, the random vector $(X_k, Y_k, X_{k'}, Y_{k'})$ admits a probability density function $f_{(X_k, Y_k, X_{k'}, Y_{k'})}$, such that $\sup_{|k-k'| \geq 1} \sup_{(s,t) \in \mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}} \int_{\mathbb{R}^{d'}} \left| G_{k,k'}(s, u, t, v) \right| du dv < \infty$, where $G_{k,k'}(\cdot, \cdot, \cdot, \cdot) = f_{(X_k, Y_k, X_{k'}, Y_{k'})}(\cdot, \cdot, \cdot, \cdot) - f_{(X,Y)}(\cdot, \cdot) f_{(X,Y)}(\cdot, \cdot)$.

Assumptions **H.1** and **H.3** are classical in a nonparametric estimation field and they are similar to those classically used in the nonrecursive case. The former is satisfied by Gaussian and Eipanechnikov kernels, while the latter is checked by linear processes, as soon as f is bounded. Note that **H.1**(i)–(ii) are technical conditions, the first allows the cancellation of the first-order term of Taylor development in the computation of the bias term, while the latter ensures the existence of the second-order term. Much more should be said about assumption **H.2**. It is particular to the recursive problem and is clearly unrestrictive, since the choice $h_n = C_n n^{-\nu}$, with $C_n \downarrow c > 0$, and $0 < \nu < 1$ is a typical example of bandwidth satisfying **H.2.** Concerning H.4, the condition **H.4(ii)** is clearly checked if m is a bounded function, and implies that

$$\mathrm{E}\left(\max_{1 \leq i \leq n} |m(Y_i)|^p\right) = O\left((\ln n)^{p/\theta}\right), \text{ for all } p \geq 1, n \geq 2.$$

The earlier condition was used by Bosq and Cheze-Payaud (1999) to study the mean square error of the Nadaraya-Watson estimator. Assumption **H.4(iii)** was used by Roussas and Tran (1992) to study the asymptotic normality of $r_n^{DW}(x)$.

Now, we can provide the main result.

[1] If a_n and b_n are two real sequences, $a_n \sim b_n$ means that the ratio a_n/b_n converges 1

Author's personal copy

2.2 Main result

Let us set

$$B_n = h_n^2 \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} \frac{1}{2} \sum_{1\le i,j\le d} \left(\frac{\partial^2 r(x)}{\partial x_i \partial x_j} + 2 \frac{\partial \ln f(x)}{\partial x_i} \frac{\partial r(x)}{\partial x_j} \right) \int_{\mathbb{R}^d} v_i v_j K(v) dv.$$

The pointwise asymptotic gaussian distribution for our class of nonparametric recursive regression estimate is given in Theorem 2.1 below, and will be proved in Sect. 3.

Theorem 2.1 *When assumptions **H.1** − **H.4**, hold, if for all* $p > 0$, $(\ln n)^{\frac{1}{\theta}} h_n^p \to 0$, *as* $n \to \infty$, *then*

$$\sqrt{nh_n^d}\left[r_n^\ell(x) - r(x) - B_n\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left[0, \frac{\sigma_\ell^2(x) V(x)}{f^2(x)}\right], \text{ as } n \to \infty,$$

for all x *such that* $f(x) > 0$, *where*

$$\sigma_\ell^2(x) = \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} f(x) \int_{\mathbb{R}^d} K^2(x) dx \text{ and } V(x) = E\left[m^2(Y_0)|X_0 = x\right] - r^2(x).$$

One may derive a simpler version of Theorem 2.1 by using an additional assumption that allows the cancellation of the bias term B_n.

Corollary 1 *Under assumptions **H.1**–**H.4** and if* $nh_n^{d+4} \to 0$ *as* $n \to 0$, *then*

$$\sqrt{nh_n^d}\left[r_n^\ell(x) - r(x)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left[0, \frac{\sigma_\ell^2(x) V(x)}{f^2(x)}\right], \text{ as } n \to \infty,$$

for all x *such that* $f(x) > 0$.

Corollary 1 is an extension of the Rousssas and Tran's (1992) result on Devroye–Wagner estimate to the general family of recursive estimators $r_n^\ell(x)$ for which the Devroye–Wagner estimate is a special case. The condition $nh_n^{d+4} \to 0$ as $n \to 0$, implies that $(\ln n)^{\frac{1}{\theta}} h_n^p \to 0$, for all $p > 0$, and satisfied by the choice $h_n = C_n n^{-\nu}$, with $C_n \downarrow c > 0$ and $1/(d+4) < \nu < 1/(d+2)$. Let us mention that **H.2(iii)** will play a key role in our methodology, in particular when we prove the negligibility of some covariance terms for $0 \le \ell \le (d-2)^+/2$, but is not necessary if $\ell > 1/2$. Also if $\ell > 1/2$, our results can be established for $\rho > 2$. So, we observe that the estimators built with 'small' values of ℓ allow some restrictions on the smooth parameter h_n and the strong mixing coefficient. However, as shown in Amiri (2009), these estimators are preferable than those built with 'large' ℓ in terms of small variance criterion.

In practice, the constants of variance appearing in Theorem 2.1 need to be estimated. To this end, one may consider using the simple Gaussian kernel and replace $f(x)$ by $f_n^\ell(x)$. There are many possibilities for constructing a consistent conditional variance estimate. One may use the functional kernel regression technique.

In order to prove Theorem 2.1, let us consider using the following decomposition.

$$r_n^\ell(x) - r(x) = \left[\tilde{r}_n^\ell(x) - r(x)\right] + \left[r_n^\ell(x) - \tilde{r}_n^\ell(x)\right],$$

Author's personal copy

where $\tilde{r}_n^\ell(x) = \tilde{\varphi}_n^\ell(x)/f_n^\ell(x)$, $\tilde{\varphi}_n^\ell(x)$ being a truncated version of $\varphi_n^\ell(x)$ defined by

$$\tilde{\varphi}_n^\ell(x) = \frac{1}{\sum_{i=1}^{n} h_i^{d(1-\ell)}} \sum_{i=1}^{n} \frac{Y_i}{h_i^{d\ell}} \mathbf{1}_{\{|Y_i| \le b_n\}} K\left(\frac{x - X_i}{h_i}\right),$$

with b_n, a sequence of real numbers which goes to $+\infty$ as $n \to \infty$. We then need the following preliminary lemmas.

Lemma 2.2 *When assumptions **H.1** and **H.2** hold, then for all* $\ell \in [0, 1]$

(a)

$$h_n^{-4}\left[Ef_n^\ell(x) - f(x)\right]^2 \longrightarrow \left[\frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}}\right]^2 b_f^2(x) \text{ as } n \to \infty;$$

(b)

$$h_n^{-4}\left[E\varphi_n^\ell(x) - \varphi(x)\right]^2 \longrightarrow \left[\frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}}\right]^2 b_\varphi^2(x) \text{ as } n \to \infty,$$

where, if $h \in C_d^2(b)$, *we set*

$$b_h(x) := \frac{1}{2} \sum_{1 \le i, j \le d} \frac{\partial^2 h}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v) dv;$$

(c) Moreover if **H.3** *holds, then*

$$nh_n^d Var f_n^\ell(x) \longrightarrow \sigma_\ell^2(x), \text{ as } n \to \infty,$$

for all x *such that* $f(x) > 0$.

Proof The results (a) and (c) of Lemma 2.2 are obtained in Amiri (2009), while (b) can be established in the same manner as (a) by substituting f with φ.

Lemma 2.3 *When assumptions **H.1**–**H.4** hold, then for all* $\ell \in [0, 1]$

(a)

$$nh_n^d Var\tilde{\varphi}_n^\ell(x) \longrightarrow \sigma_\ell^2(x)\left[r^2(x) + V(x)\right], \text{ as } n \to \infty;$$

(b)

$$nh_n^d Cov\left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x)\right] \to \sigma_\ell^2(x) r(x) \text{ as } n \to \infty.$$

Proof (a) Let us set

$$V_n^* = \sum_{k=1}^{n} \mathrm{E}Z_{k,n}^{*2} \text{ where } Z_{i,n}^* = W_{n,i} - \mathrm{E}W_{n,i}, \text{ with } W_{n,i} := \frac{K\left(\frac{x - X_i}{h_i}\right) m(Y_i) \mathbf{1}_{\{|m(Y_i)| \le b_n\}}}{h_i^{d\ell}}.$$

The variance of $\tilde{\varphi}_n^\ell(x)$ can be decomposed in variance and covariance terms as

$$\mathrm{Var}\tilde{\varphi}_n^\ell(x) = \frac{1}{n^2 h_n^{2d(1-\ell)} B_{n,d(1-\ell)}^2}\left[V_n^* + \sum_{k=1}^{n} \sum_{\substack{k'=1 \\ k \neq k'}}^{n} \mathrm{Cov}\left(Z_{k,n}, Z_{k',n}\right)\right].$$

Author's personal copy

Concerning the variance term one may write

$$\frac{V_n^*}{nh_n^{d(1-2\ell)}B_{n,d(1-2\ell)}} = \frac{nh_n^d}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2}\sum_{k=1}^n \left\{h_k^{-2d\ell}\mathrm{E}K^2\left(\frac{x-X_0}{h_k}\right)m^2(Y_0)\right.$$
$$- h_k^{-2d\ell}\mathrm{E}K^2\left(\frac{x-X_0}{h_k}\right)m^2(Y_0)\mathbf{1}_{\{|m(Y_0)|>b_n\}}$$
$$\left. - \mathrm{E}^2K\left(\frac{x-X_0}{h_k}\right)m(Y_0)\mathbf{1}_{\{|m(Y_i)|\le b_n\}}\right\} =: D_1 + D_2 + D_3.$$

Assumptions **H.4**(ii), (iii), the dominated convergence theorem and Bochner's lemma imply that

$$\int_{\mathbb{R}^d}\frac{1}{h_k^d}K^2\left(\frac{x-u}{h_k}\right)\left[V(u)+r^2(u)\right]f(u)du \to f(x)\left[V(x)+r^2(x)\right]\|K\|_2^2, \text{ as } k\to\infty.$$

On account of the above, assumption **H.2**(ii) and the Toeplitz lemma allow to deduce that

$$D_1 = \frac{nh_n^d\sum_{k=1}^n\left[h_k^{d(1-2\ell)}\int_{\mathbb{R}^d}\frac{1}{h_k^d}K^2\left(\frac{x-u}{h_k}\right)\left[V(u)+r^2(u)\right]f(u)du\right]}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2}$$
$$\to \sigma_\ell^2(x)\left[V(x)+r^2(x)\right],$$

as $n\to\infty$. Concerning the term D_2, if $b_n = (\delta\ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$, then using assumptions **H.2(ii)** and **H.4(ii)**, with the help of Markov's inequality, we have

$$|D_2| \le \frac{\|K\|_\infty^2\left\{\mathrm{E}m^4(Y_0)P\left(|m(Y_0)|>b_n\right)\right\}^{\frac{1}{2}} nh_n^d\sum_{k=1}^n h_k^{-2d\ell}}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2}$$
$$\le \frac{\|K\|_\infty^2\left\{\mathrm{E}m^4(Y_0)P\left(|m(Y_0)|>b_n\right)\right\}^{\frac{1}{2}} B_{n,-2d\ell}}{h_n^d B_{n,d(1-\ell)}^2}$$
$$= O\left[\frac{\exp\left(-\frac{\lambda b_n^\theta}{2}\right)(\ln n)^{\frac{2}{\theta}}B_{n,-2d\ell}}{h_n^d B_{n,d(1-\ell)}^2}\right]\to 0, \text{ as } n\to\infty.$$

Next for the last term D_3, from **H.2(i)-(iii)** and the logarithmic choice of b_n, one may write

$$|D_3| \le \frac{b_n^2 nh_n^d}{n^2h_n^{2d(1-\ell)}B_{n,d(1-\ell)}^2}\sum_{k=1}^n h_k^{-2d\ell}\left(\mathrm{E}K\left(\frac{x-X_i}{h_i}\right)\right)^2 = O\left(h_nb_n^2\right)\to 0, \text{ as } n\to+\infty.$$

Therefore

$$V_n^* \sim nh_n^{d(1-2\ell)}\beta_{d(1-2\ell)}f(x)\left[V(x)+r^2(x)\right]\int_{\mathbb{R}^d}K^2(u)du, \text{ as } n\to\infty.$$

Author's personal copy

It follows that

$$\frac{nh_n^d V_n^*}{n^2 h_n^{2d(1-\ell)} B_{n,d(1-\ell)}^2} \to \frac{\beta_{d(1-2\ell)} f(x)\left[V(x)+r^2(x)\right]}{\beta_{d(1-\ell)}^2} \int_{\mathbb{R}^d} K^2(u)du, \text{ as } n \to \infty.$$

Now, let us show that the covariance term of Var $\tilde{\varphi}_n^\ell(x)$ is negligible. To this end, define a sequence c_n of a real numbers tending to infinity as n goes to infinity, and write

$$\begin{aligned}
&\frac{\sum_{k=1}^{n}\sum_{k\neq k'}\sum_{k'=1}^{n} \operatorname{Cov}\left(Z_{k,n}, Z_{k',n}\right)}{n^2 h_n^{2d(1-\ell)} B_{n,d(1-\ell)}^2} \\
&\le \frac{2\left(\sum_{i=1}^{n}\sum_{i>j}\sum_{j=1}^{n} \left|A_{i,j}\right| 1_{\{1\le i-j\le c_n\}} + \sum_{i=1}^{n}\sum_{i>j}\sum_{j=1}^{n} \left|A_{i,j}\right| 1_{\{c_n+1\le i-j\le n-1\}}\right)}{\left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right)^2} \\
&\le \frac{2\left(\sum_{i=1}^{c_n}\sum_{p=1}^{n} A_{i+p,p} + \sum_{i=c_{n+1}}^{n-1}\sum_{p=1}^{n} A_{i+p,p}\right)}{\left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right)^2} := L_1 + L_2,
\end{aligned}$$

where

$$A_{i+p,p} = \frac{\left|\operatorname{Cov}\left[K\left(\frac{x-X_{i+p}}{h_{i+p}}\right) m\left(Y_{i+p}\right) \mathbf{1}_{\{|m(Y_{i+p})|\le b_n\}}, K\left(\frac{x-X_p}{h_p}\right) m\left(Y_p\right) \mathbf{1}_{\{|m(Y_p)|\le b_n\}}\right]\right|}{h_{i+p}^{d\ell} h_p^{d\ell}}.$$

On one hand, the Billingsley inequality (see e.g., Bosq and Blanke 2007) implies that

$$A_{i+p,p} \le 4b_n^2 \alpha_X(k) \left\|K\right\|_\infty^2 h_{i+p}^{-d\ell} h_p^{-d\ell},$$

and then, it follows from assumptions **H2(ii)** and **H.4(iv)** that

$$\begin{aligned}
L_2 &\le \frac{8b_n^2 \sum_{k=c_n+1}^{n-1}\sum_{p=1}^{n} \alpha_X(k) h_{p+k}^{-d\ell} h_p^{-d\ell}}{\left\|K\right\|_\infty^2 \left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right)^2} \le \frac{8b_n^2 \gamma \left\|K\right\|_\infty^2 \sum_{k=c_n}^{n-1}\sum_{p=1}^{n} k^{-\rho} h_{p+k}^{-d\ell} h_p^{-d\ell}}{\left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right)^2} \\
&\le \frac{8b_n^2 \gamma \left\|K\right\|_\infty^2 \frac{h_n^{-2d\ell} c_n^{-\rho+1}}{\rho-1} \sum_{p=1}^{n}\left(\frac{h_p}{h_n}\right)^{-d\ell}}{\left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right)^2} \\
&\le \frac{8b_n^2 \gamma \left\|K\right\|_\infty^2 c_n^{1-\rho} B_{n,-d\ell}}{nh_n^{2d} B_{n,d(1-\ell)}^2 (\rho-1)}.
\end{aligned}$$

Hence

$$nh_n^d L_2 = O\left(b_n^2 c_n^{1-\rho} h_n^{-d}\right).$$

Author's personal copy

On the other hand, regarding L_1, one has

$$A_{i+p,p} = \left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^{d'}} \int_{\mathbb{R}^{d'}} K\left(\frac{x-s}{h_{i+p}}\right) K\left(\frac{x-t}{h_p}\right) \right.$$
$$\left. \times \frac{m(u)\mathbf{1}_{\{|m(u)|\le b_n\}} m(v)\mathbf{1}_{\{|m(v)|\le b_n\}} G_{i+p,p}(s,u,t,v)}{\left(h_{i+p}h_p\right)^{d\ell}\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2} dsdtdudv \right|$$
$$\le \frac{b_n^2\left(h_{k+p}h_p\right)^{d(1-\ell)} \sup_{|k-k'|\ge 1} \sup_{(s,t)\in\mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}}\int_{\mathbb{R}^{d'}} \left|G_{k,k'}(s,u,t,v)\right| dudv}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2}.$$

Then

$$L_1 \le \frac{2b_n^2 \sum_{k=1}^{c_n}\sum_{p=1}^{n-k} h_{p+k}^{d(1-\ell)} h_p^{d(1-\ell)} \sup_{|k-k'|\ge 1} \sup_{(s,t)\in\mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}}\int_{\mathbb{R}^{d'}} \left|G_{k,k'}(s,u,t,v)\right| dudv}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2}$$
$$\le \frac{2b_n^2 c_n \sum_{p=1}^{n} h_p^{2d(1-\ell)} \sup_{|k-k'|\ge 1} \sup_{(s,t)\in\mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}}\int_{\mathbb{R}^{d'}} \left|G_{k,k'}(s,u,t,v)\right| dudv}{\left(\sum_{i=1}^n h_i^{d(1-\ell)}\right)^2}. \tag{2}$$

At this point, two cases can be distinguished according to small and large values of ℓ.

- If $\ell \in \left[\left(\frac{d-2}{2d}\right)^+, 1\right]$, then $2d(1-\ell) \le d+2$ implies $B_{n,2d(1-\ell)} \to \beta_{2d(1-\ell)} < \infty$, as $n \to \infty$, because of **H.2**(ii). It follows that

$$L_1 \le \frac{2b_n^2 c_n B_{n,2d(1-\ell)} \sup_{|k-k'|\ge 1} \sup_{(s,t)\in\mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}}\int_{\mathbb{R}^{d'}} \left|G_{k,k'}(s,u,t,v)\right| dudv}{nB_{n,d(1-\ell)}^2},$$

  which implies that

$$nh_n^d L_1 = O\left(b_n^2 c_n h_n^d\right).$$

  Thus, when $c_n := \left\lfloor h_n^{-\frac{2d}{\rho}} \right\rfloor$, and $b_n = (\delta \ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$, then

$$\frac{nh_n^d}{n^2 h_n^{2d(1-\ell)} B_{n,d(1-\ell)}^2} \sum_{k=1}^n \sum_{\substack{k'=1 \\ k\neq k'}}^n \operatorname{Cov}\left(Z_{k,n}, Z_{k',n}\right) = O\left(b_n^2 h_n^{-\frac{d(2-\rho)}{\rho}}\right) \to 0, \text{ as } n\to\infty,$$

  since $\rho > 2$.
- If $d \ge 3$, $\ell \in \left[0, \frac{d-2}{2d}\right[$, then the term L_1 cannot be studied as previously, because assumption **H.2(ii)** is not satisfied, since $2d(1-\ell) > d+2$. In this case, let us consider relation (2) and choose a real number ξ such that $\frac{1}{\rho-1} < \xi \le \frac{2}{d}$. Let us

Author's personal copy

mention that ξ exists only if $\rho > \dfrac{d+2}{2}$. Thus, we have the relation $d(\xi+1) \leq d+2$, which implies that $B_{n,d(\xi+1)} \to \beta_{d(\xi+1)} < \infty$, as $n \to \infty$, by vertue of **H.2**(ii). Also, since h_n decreases one has $\sum_{i=1}^{n} h_i^{d(1-\ell)} \geq h_1^{-d\ell} \sum_{i=1}^{n} h_i^d$. It follows that

$$\frac{c_n b_n^2 \sum_{p=1}^{n} h_p^{2d(1-\ell)}}{\left(\sum_{i=1}^{n} h_i^{d(1-\ell)}\right)^2} \leq \frac{c_n b_n^2 h_1^{d(1-\xi-2\ell)} n h_n^{d(\xi+1)} B_{n,d(\xi+1)}}{n^2 h_1^{-2d\ell} h_n^{2d} B_{n,d}^2} \leq \frac{c_n b_n^2 h_1^{d(1-\xi)} h_n^{d\xi} B_{n,d(\xi+1)}}{n h_n^d B_{n,d}^2},$$

because $0 \leq \ell < \frac{d-2}{2d} \Rightarrow 1 - \xi - 2\ell > 0,$ as long as $\xi \leq \frac{2}{d}$. Therefore, from (2) we have

$$n h_n^d L_1 = O\left(c_n b_n^2 h_n^{d\xi}\right).$$

The choices $c_n := \left\lfloor h_n^{-\frac{d(\xi+1)}{\rho}} \right\rfloor$ and $b_n = (\delta \ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$ imply the negligibility of the covariance term.

(b) Let us consider the decomposition

$$\mathrm{Cov}\left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x)\right] = \left[\sum_{i=1}^{n} h_i^{d(1-\ell)}\right]^{-2} \left[\sum_{i=1}^{n} A_{ii} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i\neq j}}^{n} A_{ij}\right] := F_1 + F_2.$$

where, for all integers s, t

$$A_{s,t} := \mathrm{Cov}\left[\frac{1}{h_s^{d\ell}} K\left(\frac{x - X_s}{h_s}\right), \frac{m(Y_t)}{h_t^{d\ell}} \mathbf{1}_{\{|m(Y_t)\leq b_n|\}} K\left(\frac{x - X_t}{h_t}\right)\right].$$

Finally, we proceed as in the proof of (a) and find

$$n h_n^d F_1 \to \sigma_\ell^2(x) r(x), \text{ and } n h_n^d F_2 \to 0, \text{ as } n \to \infty.$$

3 Proof of main result

Proof To prove the main result, we show that the asymptotic distribution of the principal term $\left[\tilde{r}_n^\ell(x) - r(x)\right]$ is normal, while the residual term $\left[r_n^\ell(x) - \tilde{r}_n^\ell(x)\right]$ is negligible. First, observe that if $b_n = (\delta \ln n)^{\frac{1}{\theta}}$ with$\delta > \frac{2}{\lambda}$, then for all $\varepsilon > 0$, we have

$$P\left(\left|\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)\right| > \varepsilon/\sqrt{n h_n^d}\right) \leq P\left(\bigcup_{i=1}^{n} \{|Y_i| > b_n\}\right)$$
$$\leq n P\left(|Y_0| > b_n\right) \leq \mathrm{E} e^{\lambda |m(Y_0)|^\theta} n^{1-\lambda\delta}.$$

So, for all $\varepsilon > 0$, $\sum_{n=1}^{\infty} P\left(\left|\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)\right| > \varepsilon/\sqrt{n h_n^d}\right) < \infty$, and the Borel-Cantelli lemma implies that

$$\sqrt{n h_n^d}\left[r_n^\ell(x) - \tilde{r}_n^\ell(x)\right] \to 0 \text{ a.s}, \quad \text{as } n \to \infty.$$

Author's personal copy

One may prove in the same manner that $f_n^\ell(x) \to f(x)$ a.s as $n \to \infty$. Next, we need to show that

$$\sqrt{nh_n^d}\left[\tilde{r}_n^\ell(x) - B_n - r(x)\right] \xrightarrow{\mathcal{L}} \mathcal{N}\left[0, \frac{\beta_{d(1-2\ell)}\|K\|_2^2 V(x)}{\beta_{d(1-\ell)}^2 f(x)}\right],$$

as $n \to \infty$. To this end, we use the following representation

$$\tilde{r}_n^\ell(x) - r(x) - B_n = \frac{1}{f_n^\ell(x)\mathrm{E} f_n^\ell(x)}\begin{bmatrix}\mathrm{E} f_n^\ell(x)\\ -\mathrm{E}\tilde{\varphi}_n^\ell(x)\end{bmatrix}^T\begin{bmatrix}\tilde{\varphi}_n^\ell(x) - \mathrm{E}\tilde{\varphi}_n^\ell(x)\\ f_n^\ell(x) - \mathrm{E} f_n^\ell(x)\end{bmatrix} + o\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

Now, applying the Cramer–Wold device and remembering that $f_n^\ell(x) \xrightarrow{a.s} f(x)$, and $\mathrm{E} f_n^\ell(x) \to f(x)$, as $n \to \infty$, the proof of Theorem 2.1 is straightforward from the following claim:

$$\sqrt{nh_n^d}\begin{bmatrix}\tilde{\varphi}_n^\ell(x) - \mathrm{E}\tilde{\varphi}_n^\ell(x)\\ f_n^\ell(x) - \mathrm{E} f_n^\ell(x)\end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_2\left\{0, \sigma_\ell^2(x)\begin{bmatrix}V(x) + r^2(x) & r(x)\\ r(x) & 1\end{bmatrix}\right\}, \text{ as } n \to \infty.$$

This last convergence is equivalent to

$$\sqrt{nh_n^d}\left\{\lambda_1\left[f_n^\ell(x) - \mathrm{E} f_n^\ell(x)\right] + \lambda_2\left[\tilde{\varphi}_n^\ell(x) - \mathrm{E}\tilde{\varphi}_n^\ell(x)\right]\right\} \xrightarrow{\mathcal{L}} \mathcal{N}\left[0, \Sigma_\ell^2(x)\right], \text{ as } n \to \infty, \quad (3)$$

for each $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1 + \lambda_2 \neq 0$, where $\Sigma_\ell^2(x) := \sigma_\ell^2(x)\left\{\lambda_1^2 + 2\lambda_1\lambda_2 r(x) + \lambda_2^2\left[V(x) + r^2(x)\right]\right\}$.

Hence, the main result will be completely proven if (3) is established. To this end, let us set

$$\tilde{\Psi}_{nj} := \lambda_1\Psi_{nj} + \lambda_2\Psi'_{nj},$$

where $\Psi_{nj} := \left[\frac{h_n^{d(2\ell-1)}}{n}\right]^{\frac{1}{2}} \frac{h_j^{-d\ell}}{B_{n,d(1-\ell)}}\left(V_{nj} - \mathrm{E}V_{nj}\right)$ and $\Psi'_{nj} := \left[\frac{h_n^{d(2\ell-1)}}{n}\right]^{\frac{1}{2}} \frac{h_j^{-d\ell}}{B_{n,d(1-\ell)}}\left(W_{nj} - \mathrm{E}.W_{nj}\right)$ with

$$V_{nj} := K\left(\frac{x - X_j}{h_j}\right) \text{ and } W_{nj} := K\left(\frac{x - X_j}{h_j}\right) m\left(Y_j\right)\mathbf{1}_{\{|m(Y_j)|\le b_n\}}.$$

Next, consider the sequences ς_n, τ_n, and r_n defined as

$$\tau_n := \lfloor \tau_0 \log n \rfloor, \quad \varsigma_n := \left\lfloor \frac{\tau_0\sqrt{nh_n^d}}{(\log n)^{\varsigma_0}} \right\rfloor \text{ and } r_n := \left\lfloor \frac{n}{\varsigma_n + \tau_n} \right\rfloor, \text{ with } \tau_0, \varsigma_0 > 0.$$

To establish (3), we use the classical Doob (1953) methodology, which consists of splitting the term

$$\sqrt{nh_n^d}\left\{\lambda_1\left[f_n^\ell(x) - f(x)\right] + \lambda_2\left[\tilde{\varphi}_n^\ell(x) - \varphi(x)\right]\right\}$$

into large blocks separated by small blocks defined by

$$T_{nm} = \sum_{j=k_m}^{k_m+\varsigma_n-1} \tilde{\Psi}_{nj} \text{ (large blocks)}, \quad T'_{nm} = \sum_{j=l_m}^{l_m+\tau_n-1} \tilde{\Psi}_{nj} \text{ (small blocks)},$$

$$T'_{nr_n+1} = \sum_{j=\bar{N}+1}^{n} \tilde{\Psi}_{nj} \text{ (rest of term)},$$

Author's personal copy

where $\bar{N} := r_n(\tau_n + \varsigma_n)$, and for $m = 1, \ldots, r_n, k_m := (m-1)(\varsigma_n + \tau_n) + 1,\ \ l_m := (m-1)(\varsigma_n + \tau_n) + \varsigma_n + 1$.

Next, let us define the partial sums $S_{n1} = \sum_{m=1}^{r_n} T_{nm},\ S_{n2} = \sum_{m=1}^{r_n} T'_{nm}$ and $S_{n3} = T'_{nr_n+1}$. Thus, we can write

$$\sqrt{nh_n^d}\left\{\lambda_1\left[f_n^\ell(x) - f(x)\right] + \lambda_2\left[\tilde{\varphi}_n^\ell(x) - \varphi(x)\right]\right\} = S_{n1} + S_{n2} + S_{n3}.$$

The goal is to prove that $\mathrm{E}S_{n2}^2$ and $\mathrm{E}S_{n3}^2$ converge to zero, while the asymptotic distribution of S_{n1} is normal. First, observe that

$$\begin{aligned}\mathrm{E}S_{n2}^2 &= \sum_{m=1}^{r_n}\mathrm{Var}(T'_{nm}) + 2\sum_{1\le i<j\le r_n}\mathrm{Cov}(T'_{ni}, T'_{nj})\\ &= \sum_{m=1}^{r_n}\sum_{i=l_m}^{l_m+\tau_n-1}\mathrm{Var}\tilde{\Psi}_{ni} + 2\sum_{m=1}^{r_n}\sum_{l_m\le i<j\le l_m+\tau_n-1}\mathrm{Cov}\left(\tilde{\Psi}_{ni}, \tilde{\Psi}_{nj}\right)\\ &\quad + 2\sum_{1\le i<j\le r_n}\sum_{s=l_i}^{l_i+\tau_n-1}\sum_{t=l_j}^{l_j+\tau_n-1}\mathrm{Cov}\left(\tilde{\Psi}_{ns}, \tilde{\Psi}_{nt}\right) := \Delta_1 + \Delta_2 + \Delta_3. \end{aligned} \tag{4}$$

The first term in (4), is decomposed as

$$\Delta_1 = \sum_{m=1}^{r_n}\sum_{i=l_m}^{l_m+\tau_n-1}\left[\lambda_1^2\mathrm{Var}\Psi_{ni} + \lambda_2^2\mathrm{Var}\Psi'_{ni} + 2\lambda_1\lambda_2\mathrm{Cov}\left(\Psi_{ni}, \Psi'_{ni}\right)\right] := \Delta_{11} + \Delta_{12} + \Delta_{13}.$$

Since h_n decreases, the choice of $b_n = (\delta \ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$, and $\theta > 1/\varsigma_0$ with the help of **H2**(**iii**) implies that

$$\begin{aligned}\Delta_{11} + \Delta_{12} &= \frac{h_n^{d(2\ell-1)}}{nB_{n,d(1-\ell)}^2}\sum_{m=1}^{r_n}\sum_{j=l_m}^{l_m+\tau_n-1} h_j^{-2d\ell}\left[\lambda_1^2\mathrm{Var}K\left(\frac{x-X_j}{h_j}\right)\right.\\ &\quad\left. + \lambda_2^2\mathrm{Var}K\left(\frac{x-X_j}{h_j}\right)Y_j\mathbf{1}_{\{|m(Y_j)|\le b_n\}}\right]\\ &\le \frac{r_n\tau_n\left(1+b_n^2\right)\|K\|_\infty^2\max(\lambda_1^2, \lambda_2^2)}{nh_n^dB_{n,d(1-\ell)}^2} \to 0,\ \text{as } n\to\infty.\end{aligned}$$

Similarly, we have $\Delta_{13} \le \frac{2\lambda_1\lambda_2 b_n r_n\tau_n\|K\|_\infty^2}{nh_n^dB_{n,d(1-\ell)}^2} \to 0$, as $n \to \infty$. In the same manner, and by also using the Cauchy-Schwartz's inequality, we get $\Delta_2 \le \frac{r_n\tau_n^2(1+b_n)^2\|K\|_\infty^2\max(\lambda_1^2,\lambda_2^2)}{nh_n^dB_{n,d(1-\ell)}^2} \to 0$, as $n \to \infty$. The last term in (4) is bounded by Billingsley inequality with the help of assumptions **H2**(**iii**) and **H.3**(i), as follows.

Author's personal copy

$$\Delta_3 = 2\sum_{1\le i<j\le r_n}\sum_{s=l_i}^{l_i+\tau_n-1}\sum_{t=l_j}^{l_j+\tau_n-1}\Big\{\lambda_1^2\mathrm{Cov}\left(\Psi_{ns},\Psi_{nt}\right)+\lambda_2^2\mathrm{Cov}\left(\Psi'_{ns},\Psi'_{nt}\right)$$
$$+\lambda_1\lambda_2\left[\mathrm{Cov}\left(\Psi_{ns},\Psi'_{nt}\right)+\mathrm{Cov}\left(\Psi_{nt},\Psi'_{ns}\right)\right]\Big\}$$
$$\le \frac{2\,(1+b_n)^2\,\|K\|_\infty^2\max(\lambda_1^2,\lambda_2^2)h_n^{d(2\ell-1)}}{nB_{n,d(1-\ell)}^2}\sum_{k=1}^{r_n-1}\sum_{j=1}^{r_n}\sum_{s=l_j}^{l_j+\tau_n-1}\sum_{t=l_j}^{l_j+\tau_n-1}(h_sh_t)^{-d\ell}\alpha_X\left[k\left(\varsigma_n+\tau_n\right)\right]$$
$$\le \frac{2\gamma\,(1+b_n)^2\,\|K\|_\infty^2\max(\lambda_1^2,\lambda_2^2)r_n\tau_n^2}{nh_n^dB_{n,d(1-\ell)}^2}\sum_{k=1}^{r_n-1}e^{-\rho k\tau_n}.$$

Therefore,

$$\Delta_3 = O\left\{\frac{b_n^2r_n\tau_n^2e^{-\rho\tau_n}}{nh_n^dB_{n,d(1-\ell)}^2}\left[1-e^{-\rho\tau_n(r_n-1)}\right]\right\}\to 0,\ \text{ as } n\to\infty,$$

as long as $b_n = (\delta\ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$, and $\theta > 1/\varsigma_0$. Now, let us prove that $\mathrm{E}S_{n3}^2 \to 0$ as $n \to 0$. One has

$$\mathrm{E}S_{n3}^2 = \sum_{j=\bar N+1}^{n}\mathrm{Var}\tilde\Psi_{nj} + 2\sum_{\bar N+1\le i<j\le n}\mathrm{Cov}(\tilde\Psi_{ni},\tilde\Psi_{nj}) := \Theta_{n1}+\Theta_{n2}. \tag{5}$$

The variance term Θ_{n1} may be written as

$$\Theta_{n1} = \sum_{j=\bar N+1}^{n}\left[\lambda_1^2\mathrm{Var}\Psi_{nj}+\lambda_2^2\mathrm{Var}\Psi'_{nj}+2\lambda_1\lambda_2\mathrm{Cov}\left(\Psi_{nj},\Psi'_{nj}\right)\right]:=\lambda_1^2\Theta_{n11}+\lambda_2^2\Theta_{n12}$$
$$+2\lambda_1\lambda_2\Theta_{n13}.$$

The first term on the right hand side of the preview decomposition satisfies the relation

$$nh_n^d\mathrm{Var}\,f_n^\ell(x)\sim\sum_{j=1}^{n}\mathrm{Var}(\Psi_{nj}) = \sum_{j=1}^{\bar N}\mathrm{Var}(\Psi_{nj})+\Theta_{n11}.$$

However, one may write

$$\sum_{j=1}^{\bar N}\mathrm{Var}(\Psi_{nj}) = \left(\frac{nh_n^d}{\bar Nh_{\bar N}^d}\right)\bar Nh_{\bar N}^d\ \mathrm{Var}\ f_{\bar N}^\ell(x).$$

Since $\bar N\sim n$, the condition $u_n\sim v_n$ implies $h_{u_n}\sim h_{v_n}$, which leads to $nh_n^d\sim\bar Nh_{\bar N}^d$, and this together with Lemma 2.2(c) imply that $\sum_{j=1}^{\bar N}\mathrm{Var}(\Psi_{nj})\to\sigma_\ell^2(x),$ as $n\to\infty$. It follows that $\Theta_{n11} = o(1)$, because $\sum_{j=1}^{n}\mathrm{Var}(\Psi_{nj})\to\sigma_\ell^2(x),$ as $n\to\infty$. Let us mention that if $\ell\ge 1/2$, then the condition $u_n\sim v_n$ implies $h_{u_n}\sim h_{v_n}$, is not necessary. Indeed, the variance term Θ_{n1} can be written as

$$\Theta_{n1} = \frac{B_{n,d(1-\ell)}^{-2}}{n}\sum_{i=\bar N+1}^{n}\left(\frac{h_i}{h_n}\right)^{d(1-2\ell)}h_i^{-d}\mathrm{Var}\left\{K\left(\frac{x-X_i}{h_i}\right)\left(1+m(Y_i)\mathbf{1}_{|m(Y_i)|\le b_n}\right)\right\}.$$

Author's personal copy

Since h_n is decreasing and $\ell \geq \frac{1}{2}$, then the Toeplitz lemma, with the help of assumption **H.2(ii)** and the convergence $h_i^{-d}\mathrm{Var}K\left(\frac{x-X_i}{h_i}\right) \to f(x)\int_{\mathbb{R}^d} K^2(x)dx$, as $i \to \infty$ imply that $\Theta_{n1} \leq \frac{\mathrm{Cste}(n-\bar{N})(1+b_n)^2}{nB^2_{n,d(1-\ell)}}$. Because of $n - \bar{N} \leq \varsigma_n + \tau_n$, it follows that $\Theta_{n1} \to 0$ as $n \to \infty$, provided $b_n = (\delta \ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$, and $\theta > 1/\varsigma_0$.

Also, in the same manner, and by replacing f_n^ℓ by $\tilde{\varphi}_n^\ell$, we can deduce from Lemma 2.3(a) that $\Theta_{n12} = o(1)$. Finally, the last term Θ_{n13} is bounded similarly to the first term by using Lemma 2.3(b).

Therefore, from $\Theta_{n13} = o(1)$, it follows that $\Theta_{n1} \to 0$ as $n \to \infty$. Now, let us study the term Θ_{n2} in (5). This can be decomposed as

$$\Theta_{n2} = 2 \sum_{\bar{N}+1\leq i<j\leq n} \left[\lambda_1^2 \mathrm{Cov}\left(\Psi_{ni}, \Psi_{nj}\right) + \lambda_2^2 \mathrm{Cov}\left(\Psi'_{ni}, \Psi'_{nj}\right) + 2\lambda_1\lambda_2 \mathrm{Cov}\left(\Psi_{ni}, \Psi'_{nj}\right)\right].$$

As in the proof of Lemma 2.3, one may show that

$$\sum_{\bar{N}+1\leq i<j\leq n} \left[\lambda_1^2 \mathrm{Cov}\left(\Psi_{ni}, \Psi_{nj}\right) + \lambda_2^2 \mathrm{Cov}\left(\Psi'_{ni}, \Psi'_{nj}\right)\right] \to 0, \text{ as } \to 0$$

and

$$\sum_{\bar{N}+\leq i<j\leq n} \mathrm{Cov}\left(\Psi_{ni}, \Psi'_{nj}\right) \leq \sum_{1\leq i<j\leq n} \mathrm{Cov}\left(\Psi_{ni}, \Psi'_{nj}\right) \to 0, \text{ as } \to 0.$$

Hence, $\Theta_{n2} \to 0$, as $\to 0$. To complete the proof we must show that the asymptotic distribution of S_{n1} is normal. To this end let us check the Lindeberg-Feller conditions for S_{n1}. First, we consider a sequence of iid random variables $Z_{n1}, \ldots, Z_{nr_n}$, having the same distribution as T_{nm}. Then, $\mathrm{E}Z_{n1} = 0$ and if $\Phi_{T_{nm}}$ is the characteristic function (ch.f.) of T_{nm}, then $\Phi^{r_n}_{T_{nm}}$ is the ch.f. of the random variable $\sum_{m=1}^{r_n} Z_{nm}$. To establish the asymptotic normality of S_{n1}, it suffices to prove that the variables $\sum_{m=1}^{r_n} Z_{nm}$ and $\sum_{m=1}^{r_n} T_{nm}$ have the same distribution, and that this latter is Gaussian. By the Volkonskii and Rozanov (1959) lemma, one has

$$\left|\mathrm{E}\prod_{m=1}^{r_n} e^{itT_{nm}} - \prod_{m=1}^{r_n} \mathrm{E}e^{itT_{nm}}\right| \leq 8(r_n - 1)\alpha(\tau_n) \leq \rho_0 r_n e^{-\rho_1\tau_n} \to 0, \text{ as } n \to \infty.$$

It follows that $\left|\mathrm{E}\prod_{m=1}^{r_n} e^{itT_{nm}} - \Phi^{r_n}_{T_n}\right| \to 0$, as $n \to \infty$. Then, it suffices to prove that $\Phi^{r_n}_{T_{nm}}$ converges to the characteristic function of a Gaussian random variable. To this end, we proceed as follows. Set $Z'_{nm} := \frac{Z_{nm}}{s_n}$, where $s_n^2 := \sum_{m=1}^{r_n} \mathrm{Var}Z_{nm}$. One has $s_n^2 \to \Sigma_\ell^2(x)$, as $n \to \infty$. Indeed, $s_n^2 = \sum_{m=1}^{r_n} \mathrm{Var}T_{nm} \to \Sigma_\ell^2(x)$, as $n \to \infty$, because on one hand we have from Lemmas 2.2 and 2.3:

$$\begin{aligned}\mathrm{Var}S_{n1} &\sim nh_n^d\left\{\lambda_1^2\mathrm{Var}f_n^\ell(x)+\lambda_2^2\mathrm{Var}\tilde{\varphi}_n^\ell(x)+2\lambda_1\lambda_2\mathrm{Cov}\left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x)\right]\right\}\\ &\to \Sigma_\ell^2(x), \text{ as } n \to \infty,\end{aligned}$$

Author's personal copy

and one may show, on the other hand, as for Δ_2, that $\sum_{1\le i<j\le r_n} \mathrm{Cov}(T_{ni}, T_{nj}) \to 0$, as $n \to \infty$. Hence, the variables Z'_{nm} are iid, $\mathrm{E}Z'_{n1} = 0$ and $\sum_{m=1}^{r_n} \mathrm{Var}Z'_{nm} = 1$. By virtue of the Lindeberg conditions (c.f. Loève 1963), we have to show that for all $\varepsilon > 0$, $\sum_{m=1}^{r_n} \mathrm{E}\left(Z'^2_{nm}\mathbf{1}_{\{|Z'_{nm}|>\varepsilon\}}\right) \to 0$, as $n \to \infty$. Noting that $|T_{nm}| \le \frac{\varsigma_n \|K\|_\infty (1+b_n)}{\sqrt{nh_n^d} B_{n,d(1-\ell)}}$, and applying Markov's inequality, one has

$$\begin{aligned}
\sum_{m=1}^{r_n} \mathrm{E}\left(Z'^2_{nm}\mathbf{1}_{\{|Z'_{nm}|>\varepsilon\}}\right) &= \sum_{m=1}^{r_n} \mathrm{E}\left(\frac{T_{nm}^2}{s_n^2}\mathbf{1}_{\{|T_{nm}|>\varepsilon s_n\}}\right)\\
&\le \frac{\varsigma_n^2 (1+b_n)^2 \|K\|_\infty^2}{nh_n^d B_{n,d(1-\ell)}^2 s_n^2} \sum_{m=1}^{r_n} P\left(|T_{nm}|>\varepsilon s_n\right)\\
&\le \left[\frac{\varsigma_n (1+b_n)}{\sqrt{nh_n^d}} \cdot \frac{\|K\|_\infty \varepsilon^{-1}}{s_n B_{n,d(1-\ell)}}\right]^2 \to 0, \text{ as } n \to \infty,
\end{aligned}$$

if $b_n = (\delta \ln n)^{\frac{1}{\theta}}$ with $\delta > \frac{2}{\lambda}$, and $\theta > 1/\varsigma_0$.

References

Ahmad I, Lin PE (1976) Nonparametric sequential estimation of a multiple regression function. Bull Math Stat 17:63–75

Amiri A (2009) Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortement mélangeant. C R Acad Sci Paris Ser I 347:309–314

Amiri A (2012) Recursive regression estimators with application to nonparametric prediction. J Nonparam Stat 24:169–186

Bosq D, Blanke D (2007) Inference and prediction in large dimensions, Wiley series in probability and statistics. Wiley, Chichester.

Bosq D, Cheze-Payaud N (1999) Optimal asymptotic quadratic error of nonparametric regression function estimates for a continuous-time process from sampled-data. Statistics 32(3):229–247

Devroye L (1981) On the almost everywhere convergence of nonparametric regression function estimates. Ann Stat 9:1310–1319

Devroye L, Wagner TJ (1980) On the L^1 convergence of kernel estimators of regression functions with application in discrimination. Z Wahrschein Verw Get 51:15–25

Doob J (1953) Stoch process. Wiley, New York

Greblecki W, Pawlak M (1987) Necessary and sufficient consistency conditions for a recursive kernel regression estimate. J Multivar Anal 23:67–76

Gyorfi L, Hardle W, Sarda P, Vieu P (1989) Nonparametric curve estimation from time Series. Springer, Lecture notes in statistics

Györfi L, Khöler M, Krzyżak A, Walk H (2002) A distribution-free theory of nonparametric regression. Springer, New york

Krzyżak A (1992) Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation. IEEE Trans Inform Theory 38:1323–1338

Loève M (1963) Probab theory. New Jersey Van Nostrand, Princeton

Prakasa-Rao BLS (1983) Nonparametric functional estimation. Academic Press, New-York

Qin YS (1995) Asymptotic distribution of a recursive kernel estimator for a nonparametric regression function under dependent sampling. Mathematica Applicata 8(1):7–13

Roussas GG (1990) Nonparametric regression estimation under mixing conditions. Stoch Process Appl 36(1):107–116

Roussas GG, Tran LT (1992) Asymptotic normality of the recursive kernel regression estimate under dependence conditions. Ann Stat 20(1):98–120

Volkonskii VA, Rozanov YuA (1959) Some limit theorems for random functions. Theory Probab Appl 4: 178–197

Author's personal copy

Walk H (2001) Strong universal pointwise consistency of recursive regression estimates. Ann Inst Stat Math 53(4):691–707

Wang L, Liang HY (2004) Strong uniform convergence of the recursive regression estimator under φ-mixing conditions. Metrika 59(3):245–261

Yoshihara K (1994) Weakly dependent stochastic sequences and their applications: curve estimation based on weakly dependent data (Vol. IV). Sanseido, Tokyo.