Nonparametric Regression Estimation under Weak Dependence beyond Mixing and Association

Patrick Ango Nze, Peter Bühlmann and Paul Doukhan
Université Lille 3, ETH Zürich and Université Cergy Pontoise

July 1999

Abstract

We consider various asymptotic aspects of the classical kernel regression estimator in the case of a very general class of stationary processes. The processes are not required to satisfy a mixing or association condition, we only assume a mild weak dependence property which has been introduced by Doukhan and Louhichi [7]. This notion of weak dependence, extending concepts like mixing or association, is broad enough to include many interesting examples such as very general Bernoulli shifts or Markovian processes.

We investigate under such a weak dependence assumption pointwise kernel estimation of conditional expectations and a local bootstrap yielding interval estimates. We justify the 'whitening by windowing principle' for kernel estimation, and thus indicate its general validity, saying that asymptotic properties remain in first order the same as for independent samples. The proofs borrow previously used strategies, but precise arguments are developed under the new aspect of general weak dependence.

Key Words and phrases. Blocking, bootstrap, central limit theorem, mixing, nonparametric estimation, positive dependence, Rosenthal inequality, stationary sequence, time series

AMS subject classifications (1991) Primary 60F05, 62M99; secondary 60E15, 60G10, 60G99, 62G07, 62G09.

1 Introduction

The analysis of statistical procedures for dependent data usually relies on some decay of dependence as the distance, say in time or space, between observations increases. The most popular notions describing such a decay of dependence are mixing and association. But such mixing or association conditions can be very hard to verify for particular models or are even too strong to be true. Not much is known in asymptotic theory about the behavior of a statistical procedure when the data generating stationary process exhibits dependencies which are beyond mixing or association. Examples where mixing and/or association fail to hold include the following: (i) Bernoulli shifts driven by discrete innovations, (ii) Markov processes driven by discrete innovations and (iii) processes arising from model- or sieve-based time series bootstraps. More precise definitions are given in section 2.2.

Doukhan and Louhichi [7] have introduced a new concept of weak dependence for stationary processes which generalizes the notions of mixing and association. Relaxation of mixing or association conditions and assuming only the new notion of weak dependence yields a fairly tractable framework for the analysis of statistical procedures with very general data generating processes, e.g. the examples (i)-(iii) mentioned above.

For a stationary time series, weak dependence as defined in [7] is measured in terms of covariances of functions. For convenient functions h and k, we assume that

$$\mathrm{Cov}(h(\text{`past'}), k(\text{`future'})) \tag{1.1}$$

is small when the distance between the 'past' and the 'future' is sufficiently large. For the functions h and k, we focus on the class of bounded Lipschitz functions and modifications thereof. This class is small enough to actually prove that for various processes, the quantities in (1.1) can be bounded uniformly over the function class as a function of the distance between the 'past' and the 'future'. On the other hand, the function class is rich enough to obtain high order moment bounds and central limit theorems for sums whenever such uniform bounds for (1.1) hold. Alternative function classes are also possible, and another similar proposal for weak dependence has been given in [2].

Under the general notion of weak dependence in terms of requiring only a suitable decay of the quantity in (1.1), it is still possible to get fairly good bounds for moment and exponential inequalities. For a real-valued, mean zero, stationary and weakly dependent process $(Z_t)_{t\in\mathbb{Z}}$, Rosenthal inequalities bound $|\mathbb{E}[(\sum_{t=1}^n Z_t)^q]|$ as $O(n^{q/2})$ for integers $q \geq 2$, but the constants in the O-term are more complicated than for mixing or associated sequences. And likewise, exponential inequalities for $\mathbb{P}[|\sum_{t=1}^n Z_t| \geq x]$ are still possible under general weak dependence but with a less tight bound than under mixing or association. Another complication arises when dealing with transformed values $g(Z_t, \ldots, Z_{t-v})$ for fixed v, where $g(.)$ is non-smooth: the quantity in (1.1) only controls explicitly for transforms $g(.)$ which are bounded Lipschitz (or modifications thereof), although non-smooth $g(.)$'s can be handled under additional concentration assumptions for the process $(Z_t)_{t\in\mathbb{Z}}$. All this indicates that some of the strategies for various proofs can be borrowed from previous works in mixing or association, but the precise arguments have to be developed under the new aspect of weak dependence.

We focus on the specific problem of point and interval estimation of conditional expectations. Consider a strictly stationary process $(Z_t)_{t\in\mathbb{Z}}$ taking values in $\mathbb{R}^D$ for some $D \in \mathbb{N}$: $(Z_t)_{t\in\mathbb{Z}}$ does not need to satisfy a mixing or association, but only a weak dependence assumption as mentioned above. For $2 \leq D = 1 + d$, writing the components as $Z_t = (X_t, Y_t)$ with $X_t \in \mathbb{R}^d$ and $Y_t \in \mathbb{R}$, the problem of interest here is nonparametric point and interval estimation of the function

$$r(x) = \mathbb{E}(Y_0 | X_0 = x).$$

For simplicity, we often restrict ourselves here to $D = 2$ so that $X_t \in \mathbb{R}$. But extensions to $D = 1 + d > 2$ with $X_t \in \mathbb{R}^d$ are straightforward. Known estimation techniques for $r(x)$, namely the kernel estimator and a local bootstrap for it, are shown to have essentially the same first order asymptotic properties as in the independent case. There seem to be little difficulties to extend our results to other smoothing techniques such as local polynomial estimators. This phenomenon is known as the 'whitening by windowing' principle (see [13]) and was first shown for mixing processes in [24]. Thus, the results here indicate validity of this principle for general windowing and smoothing methods applied to a broad class of stationary processes including non-mixing or non-associated sequences. Our strategy of proofs is based on the Lindeberg technique which seems more appropriate under general weak dependence than blocking methods.

The paper is organized as follows. The definition of weak dependence and some examples satisfying such a condition are given in section 2, section 3 describes the point and interval estimators for conditional expectations, section 4 provides asymptotic properties thereof, including a small simulation study, and section 5 contains the proofs.

2 Weak dependence

2.1 Definition

We define first the new notion of dependence, thereby closely following [7]. Generally, let E be some normed measurable space with norm $\|.\|$, although we restrict later attention to the case where $E = \mathbb{R}^D$. Denote by $\mathbb{L}^\infty(E^u)$ $(u \in \mathbb{N})$ the set of measurable bounded functions on E^u and set $\mathbb{L}^\infty = \bigcup_{u=1}^\infty \mathbb{L}^\infty(E^u)$. We then equip E^u with the l^1-norm $\|(x_1, \ldots, x_u)\|_1 = \|x_1\| + \ldots + \|x_u\|$, where $x \in E^u$. Moreover, denote by $\mathrm{Lip}(h) = \sup_{x \neq y} \frac{|h(x)-h(y)|}{\|x-y\|_1}$ the Lipschitz modulus of a function $h : E^u \to \mathbb{R}$ with respect to the l^1-norm in E^u. Define

$$\mathcal{L} = \bigcup_{u=1}^{\infty} \{h \in \mathbb{L}^\infty(E^u);\ \mathrm{Lip}(h) < \infty,\ \|h\|_\infty \leq 1\}.$$

Definition 2.1 *([7]) The E-valued sequence $(Z_t)_{t\in\mathbb{Z}}$ is called $(\theta, \mathcal{L}, \psi)$-weak dependent if for some monotone sequence $\theta = (\theta_r)_{r\in\mathbb{N}}$ decreasing to zero at infinity and some real-valued function ψ with arguments $(h, k, u, v) \in \mathcal{L}^2 \times \mathbb{N}^2$,*

$$|\mathrm{Cov}(h(Z_{i_1}, \ldots, Z_{i_u}), k(Z_{j_1}, \ldots, Z_{j_v}))| \leq \psi(h, k, u, v)\theta_r$$

for any u-tuple $(i_1, \ldots, i_u)$, any v-tuple $(j_1, \ldots, j_v)$ with $i_1 \leq \ldots \leq i_u < i_u + r \leq j_1 \leq \ldots \leq j_v$, and all $u, v \in \mathbb{N}$.

Various choices for the function ψ turn out to be convenient. Particularly, consider

$$\psi_1(h, k, u, v) = u\mathrm{Lip}(h) + v\mathrm{Lip}(k),\ \psi_1'(h, k, u, v) = v\mathrm{Lip}(k),$$
$$\psi_2(h, k, u, v) = uv\mathrm{Lip}(h)\mathrm{Lip}(k),\ \psi_2'(h, k, u, v) = v\mathrm{Lip}(h)\mathrm{Lip}(k),$$

where the functions h and k are defined on E^u and E^v, respectively $(u, v \in \mathbb{N})$. Clearly $(\theta, \mathcal{L}, \psi_j')$-weak dependence implies $(\theta, \mathcal{L}, \psi_j)$-weak dependence $(j = 1, 2)$. The functions ψ_j $(j = 1, 2)$ are adapted to provide moment bounds by using techniques in [7], whereas the functions ψ_j' $(j = 1, 2)$ provide nice CLT's via the Lindeberg-Rio method, used e.g. in [5]. The distinction with the subindex 1 and 2 corresponds to our examples, where we always consider ψ_1, ψ_1' for Bernoulli shifts and Markov processes, and ψ_2, ψ_2' for associated sequences, see section 2.2.

If the class $\mathcal{L}$ is replaced by $\mathbb{L}^\infty$ and $\psi(h,k,u,v) = 4\|h\|_\infty\|k\|_\infty$, one obtains strong mixing processes with $\theta_r = \alpha_r$ (as defined by Rosenblatt, see e.g. [6]). But such strong mixing conditions refer to σ-algebras rather than to random variables which is often an unnecessarily strong requirement. In section 2.2, we will discuss some examples where the mixing condition is too restrictive.

The notion of $\mathcal{L}$-weak dependence can also be modified to deal with indicator functions, which are not Lipschitzian, and empirical processes. In the latter case with indicators of half lines in $\mathbb{R}$, we consider instead of $\mathcal{L}$ the class,

$$\mathcal{I} = \bigcup_{u=1}^{\infty} \left\{ \bigotimes_{i=1}^{u} g_{x_i}; x_i \in \mathbb{R}^+ \text{ for } i = 1, \cdots, u \right\} \text{ where } g_x(y) = \mathbb{1}_{\{x \le y\}} - \mathbb{1}_{\{y \le -x\}}, \quad x \in \mathbb{R}^+.$$

Under $\mathcal{L}$-weak dependence and additional regularity assumptions for the underlying process, $\mathcal{I}$-weak dependence can be established. For instance, we obtain uniform covariance bounds over $\mathcal{I}$ as follows: $v\sqrt{u+v}\theta_r$, $(u+v)^{4/3}\theta_r^{1/3}$ and $(u+v)\theta_r^{1/3}$ in the $(\theta, \mathcal{L}, \psi_1')$-, $(\theta, \mathcal{L}, \psi_2)$- and $(\theta, \mathcal{L}, \psi_2')$-weak dependent cases, respectively.

2.2 Examples

We introduce here examples where weak dependence as defined in section 2.1 holds. Note that such weak dependence properties remain valid when transforming with bounded Lipschitz functions and adding independent noise. Let $(\eta_t)_{t\in\mathbb{Z}}$ be some $\mathbb{R}^{D'}$-valued i.i.d. sequence and $g : E^v \to \mathbb{R}^{D'}$ a bounded Lipschitz function. Then, if the stationary process $(Z_t)_{t\in\mathbb{Z}}$ is weakly dependent (with some ψ) and independent of $(\eta_t)_{t\in\mathbb{Z}}$, then

$$Z_t' = g(Z_{t-1}, \ldots, Z_{t-v}) + \eta_t \ (t \in \mathbb{Z})$$

has the same weak dependence properties as $(Z_t)_{t\in\mathbb{Z}}$ [only multiply ψ by some constant and replace θ_r by θ_{r-v}]. Moreover, if η_t has a bounded density, then the finite-dimensional marginal distributions of Z_t' have also a bounded density, as required later in (4.4). Such a model is often of interest in functional estimation of conditional expectations.

2.2.1 Bernoulli shifts

Definition 2.2 *Let $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of real-valued random variables and $F : \mathbb{R}^{\mathbb{Z}} \to E$ be a measurable function. The sequence $(Z_t)_{t\in\mathbb{Z}}$ defined by*

$$Z_t = F(\xi_{t-j}, j \in \mathbb{Z}), \tag{2.1}$$

is called a Bernoulli shift.

The class of Bernoulli shifts is very general. It provides examples of processes that are weakly dependent, but not mixing (see [25]).

Consider the following bound. For any $k \in \mathbb{N}$, let $\delta_k \in \mathbb{R}$ be such that

$$\sup_{t\in\mathbb{Z}} \mathbb{E}\|F(\xi_{t-j}, j \in \mathbb{Z}) - F(\xi_{t-j}\mathbb{1}_{|j|<k}, j \in \mathbb{Z})\| \le \delta_k. \tag{2.2}$$

Such sequences $(\delta_k)_{k\in\mathbb{N}}$ are related to the modulus of uniform continuity of F. The sequence $(\delta_k)_{k\in\mathbb{N}}$ can be evaluated under regularity conditions on the function F. For example, if

$$\|F(u_j : j \in \mathbb{Z}) - F(v_j : j \in \mathbb{Z})\| \le \sum_{j\in\mathbb{Z}} a_j |u_j - v_j|^b,$$

for positive constants $(a_j)_{j\in\mathbb{Z}}$, for some $0 < b \leq 1$, and if the sequence ξ_t has finite b-th order moment for all t, one can then choose $\delta_k = \sum_{|j|\geq k} a_j \mathbb{E}|\xi_j|^b$. Bernoulli shifts then satisfy the condition of $(\theta, \mathcal{L}, \psi_1)$-weak dependence with $\theta_r = \delta_{r/2}$, see Corollary 2 and Lemma 8 in [7]. Note that in case of a causal Bernoulli shift, i.e., $Z_t = F(\xi_{t-j}, j \in \mathbb{N}_0)$, then also $(\theta, \mathcal{L}, \psi_1')$-weak dependence holds with $\theta_r = \delta_r$.

Many of the statistical models may be expressed as Bernoulli shifts. For example, consider an $\mathbb{R}^d$-valued Markov process driven by the recurrence equation,

$$Z_t = f(Z_{t-1}, \xi_t) \ (t \in \mathbb{Z}) \tag{2.3}$$

for some i.i.d. sequence $(\xi_t)_{t\in\mathbb{Z}}$ with $\mathbb{E}(\xi_0) = 0$, ξ_t independent from $\{Z_s; s < t\}$ and $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ for some $d \geq 1$. Then, the function F is defined (if it exists) implicitly by the relation

$$F(x) = f(F(x'), x_0), \text{ where } x = (x_0, x_{-1}, x_{-2}, \ldots), \ x' = Bx = (x_{-1}, x_{-2}, x_{-3}, \ldots),$$

and B denotes the backshift operator.

Example 1. Consider the real-valued functional autoregressive model with representation in (2.3) as $f(u, v) = r(u) + v$, where $r : \mathbb{R} \to \mathbb{R}$ and $v \in \mathbb{R}$. In this case, if $|r(u) - r(u')| \leq c|u - u'|$ for all $u, \ u' \in \mathbb{R}$ and for some $0 \leq c < 1$, and if the innovation process $(\xi_t)_{t\in\mathbb{Z}}$ has expectation $\mathbb{E}(\xi_0) = 0$, then $\theta_r = \delta_r = Cc^r$ for some constant $C > 0$.

Remark 2.1. Let $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of i.i.d. Bernoulli variables with parameter s. The AR(1) process $(Z_t)_{t\in\mathbb{Z}}$ with innovations $(\xi_t)_{t\in\mathbb{Z}}$ and AR parameter $a \in]0, \frac{1}{2}]$ is defined by

$$Z_t = aZ_{t-1} + \xi_t = \sum_{j\geq 0} a^j \xi_{t-j}.$$

This sequence is a Bernoulli shift with $F(x) = \sum_{j\geq 0} a^j x_j$, and $\delta_k = s \sum_{j\geq k} a^j$, but it is known to be non mixing (see [25]). Note that concentration holds, e.g., Z_t is uniform if $s = \frac{1}{2}$ and it has a Cantor marginal distribution if $s = \frac{1}{3}$. Hence, without a regularity condition on the marginal distribution of ξ_0, Bernoulli shifts might not be mixing. For non-causal linear processes, we do not even know sufficient conditions mixing.

Regarding representation (2.3), we assume that Z_0 independent of the sequence $(\xi_t)_{t\in\mathbb{N}}$. Suppose that for some $0 \leq c_i < 1$, f satisfies the following condition,

$$\mathbb{E}|f(0, \xi_1)| < \infty \text{ and } \mathbb{E}|f(x, \xi_1) - f(y, \xi_1)| \leq \sum_{i=1}^{d} c_i |x_i - y_i|, \ c = \sum_{i=1}^{d} c_i < 1 \text{ for all } x, y \in \mathbb{R}^d. \tag{2.4}$$

Duflo [10] shows that under the condition in (2.4), the Markov process $(Z_t)_{t\in\mathbb{N}}$ has a stationary distribution μ with finite first moment. Assume now in addition that Z_0 is distributed with μ, i.e., the Markov chain is stationary. Then, if (2.4) holds, such a Markov chain is $(\theta, \mathcal{L}, \psi_1')$-weak dependent with $\theta_r = c^r \mathbb{E}|Z_0|$.

Example 2. Consider the real-valued functional ARCH model with representation in (2.3) as $f(u, v) = s(u)v$, where $s : \mathbb{R} \to \mathbb{R}^+$ and $v \in \mathbb{R}$. Assuming $|s(u) - s(u')| \leq c|u - u'|$ for some $0 \leq c < 1$ and for all $u, u' \in \mathbb{R}$, and if the innovation process $(\xi_t)_{t\in\mathbb{N}}$ satisfies $\mathbb{E}(\xi_1) = 0$, $\mathbb{E}|\xi_1|^2 = 1$, then (2.4) with $d = 1$ holds and implies $\theta_r = c^r \mathbb{E}|Z_0|$.

Example 3. Bootstraps for time series are often based on a Markovian model such as in (2.3). Given data $Z_1, \ldots, Z_n$ and an estimate $\hat{f}(.)$, residuals $\hat{\xi}_2, \ldots, \hat{\xi}_n$ are computed. The distribution P_ξ

of ξ_t can then be estimated by $\hat{P}_\xi = (n-1)^{-1}\sum_{t=2}^n \Delta_{\tilde{\xi}_t}$, where $\tilde{\xi}_t = \hat{\xi}_t - (n-1)^{-1}\sum_{s=2}^n \hat{\xi}_s$ are the centered residuals and Δ_x the point mass at $x \in \mathbb{R}$. The bootstrapped process is defined recursively,

$$Z_t^* = \hat{f}(Z_{t-1}^*, \xi_t^*),$$

with $(\xi_t^*)_{t\in\mathbb{N}}$ an i.i.d. sequence, ξ_t^* i.i.d $\sim \hat{P}_\xi$, independent from $\{Z_s^*; s < t\}$. It is *not* required to resample ξ_t^* from a smoothed version of $\hat{P}_\xi$. Smoothed resampling, requiring specification of a smoothing parameter, was proposed in [12] only for technical reasons ensuring mixing; and similarly in [2] in a different model for proving some weak notion of mixing. But such a regularization, which is practically very inconvenient, is not necessary: $(\theta, \mathcal{L}, \psi_1')$-weak dependence of $(Z_t^*)_{t\in\mathbb{Z}}$ with probability tending to 1, as $n \to \infty$, follows easily from some consistency result for $\hat{f}(.)$. What still remains to be shown in a case by case manner are suitable consistency properties of the bootstrap approximation for the statistic of interest, thereby using crucially a weak dependence property.

2.2.2 Chaotic representations

We now specialize the representation in (2.1) to expansions associated with the discrete chaos generated by the sequence $(\xi_t)_{t\in\mathbb{Z}}$. In a condensed formulation we write,

$$\begin{aligned} F(x) &= \sum_{k=0}^{\infty} F_k(x) \ (x \in \mathbb{R}^{\mathbb{Z}}), \\ F_k(x) &= \sum_{j_1\in\mathbb{Z}} \sum_{j_2\in\mathbb{Z}} \cdots \sum_{j_k\in\mathbb{Z}} a^{(k)}_{j_1,\ldots,j_k} x_{j_1} x_{j_2} \ldots x_{j_k} \ (k \geq 1), \end{aligned} \tag{2.5}$$

where $F_k(x)$ denotes the k-th order chaos contribution and $F_0(x) = a_0^{(0)}$ is only a centering constant. In short we write in the vectorial notation, $F_k(x) = \sum_{j\in\mathbb{Z}^k} a_j^{(k)} x_j$. Processes associated with a finite number of chaos, i.e., $F_k \equiv 0$ if $k > k_0$ for some $k_0 \in \mathbb{N}$, are also called Volterra processes. A first example of Volterra processes is the class of linear processes, including ARMA models being widely used in statistics. It corresponds to the above expansion with $F_k(x) \equiv 0$ for all $k > 1$.

A simple and general condition for $\mathbb{L}^1$-convergence of the expansion in (2.5), still written in a condensed notation, is $\sum_{k=0}^{\infty}\sum_{j\in\mathbb{Z}^k} |a_j^{(k)}| \mathbb{E}|\xi_0|^k < \infty$ which allows to define the distribution of such shift processes. $(\theta, \mathcal{L}, \psi_1)$-weak dependence holds with $\theta_r = \delta_{r/2}$ for

$$\delta_r = \sum_{k=0}^{\infty} \sum_{j\in\mathbb{Z}^k; \|j\|_\infty > r} |a_j^{(k)}| \mathbb{E}|\xi_0|^k < \infty.$$

Under causality, also $(\theta, \mathcal{L}, \psi_1')$-weak dependence holds with $\theta_r = \delta_r$.

Example 4. Consider the simple bilinear process with the recurrence equation,

$$Z_t = aZ_{t-1} + bZ_{t-1}\xi_{t-1} + \xi_t \ (t \in \mathbb{Z}).$$

Such causal processes are associated to the chaotic representation in (2.5) with

$$F(x) = \sum_{j=0}^{\infty} x_j \prod_{s=1}^{j} (a + bx_s), \ x = (x_0, x_1, x_2, \ldots).$$

If $c = \mathbb{E}|a + b\xi_0| < 1$ then $\delta_r = \theta_r = \frac{c^r(r+1)}{1-c}$ has a geometric decay rate under Tong's [27] stationarity condition and $(\theta, \mathcal{L}, \psi_1')$-weak dependence holds. More general affine models as in [15] may be considered analogously.

2.2.3 ψ_2-dependent sequences

Definition 2.3 *([11]) The sequence $(Z_t)_{t\in\mathbb{Z}}$ is associated, if for all coordinatewise increasing real-valued functions h and k,*

$$\mathrm{Cov}(h(Z_t, t\in A), k(Z_t, t\in B)) \geq 0$$

for all finite subsets A and B of $\mathbb{Z}$.

Associated sequences are $(\theta, \psi_2, \mathcal{L})$-weak dependent with $\theta_r = \sup_{k\geq r} \mathrm{Cov}(X_0, X_r)$ (see [7]). Note that broad classes of examples of associated processes result from the fact that any independent sequence is associated and that monotonicity preserves association (for this, see [18]).

The case of Gaussian sequences is analogous by setting $\theta_r = \sup_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$. Also, one may consider combinations of sums of Gaussian and associated sequences, or Bernoulli shifts driven by stationary associated sequences.

Note that for associated or Gaussian sequences, ψ_2' replaces ψ_2 if $\theta_r = \sup_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$ is replaced by $\theta_r = \sum_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$.

3 Nonparametric estimation of conditional expectations

3.1 Point estimation

We restrict ourselves in the sequel to the case of stationary processes $(Z_t)_{t\in\mathbb{Z}}$ with $Z_t = (X_t, Y_t)$ where $X_t,\ Y_t \in \mathbb{R}$. The quantity of interest is $r(x) = \mathbb{E}(Y_0|X_0 = x)$. The extension to the case where $X_t \in \mathbb{R}^d$ for some $d > 1$ is straightforward. Let K be some kernel function with integral 1, Lipschitzian and rapidly convergent to 0 at infinity. For simplicity, we assume throughout the paper that it is compactly supported. The kernel estimator (see e.g. [26]) is defined by

$$\hat{r}(x) = \hat{r}_{n,h}(x) = \frac{\hat{g}_{n,h}(x)}{\hat{f}_{n,h}(x)} \text{ if } \hat{f}_{n,h}(x) \neq 0; \qquad \hat{r}(x) = 0 \text{ otherwise,}$$

$$\hat{f}(x) = \hat{f}_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right), \qquad \hat{g}(x) = \hat{g}_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right) Y_t.$$

Here $h = (h_n)_{n\in\mathbb{N}}$ is a sequence of positive real numbers. We always assume that $h_n \to 0,\ nh_n \to \infty\ (n \to \infty)$.

3.1.1 Bias

We briefly recall the classical analysis for the deterministic part.

Definition 3.1 *Let $\rho = a + b$ with $a \in \mathbb{N}$ and $0 < b \leq 1$. Set*

$$\mathcal{C}_\rho = \Big\{u : \mathbb{R} \to \mathbb{R};\quad u \in \mathcal{C}_a \text{ and there exists } A \geq 0 \text{ such that } |u^{(a)}(x) - u^{(a)}(y)| \leq A|x - y|^b \text{ for all } x, y \text{ in any compact subset of } \mathbb{R}\Big\}.$$

Here, $\mathcal{C}_a$ is the set of a-times continuously differentiable functions and $\mathcal{C}_\rho$ is known as the set of ρ-regular functions.

Using only the stationarity assumption we have, $\mathbb{E}(\hat{g}(x)) = g_h(x)$ with $g_h(x) = \int_{-\infty}^{\infty} K(s)g(x-sh)ds$. The study of the bias $b_h(x) = g_h(x) - g(x)$ is purely analytical and does not depend on dependence properties of the sequence $(Z_t)_{t\in\mathbb{Z}}$.

Assuming $g \in \mathcal{C}_\rho$, one can choose a kernel function K of order ρ (not necessarily nonnegative, see [21] or [1]) such that the bias b_h satisfies

$$b_h(x) = g_h(x) - g(x) = O(h^\rho)$$

where the O-term is uniform on any compact subset of $\mathbb{R}$ (see e.g. [26]).

If ρ is an integer with $b = 1$, $\rho = a - 1$, then with an appropriately chosen kernel K of order ρ,

$$b_h(x) \sim \frac{\int\ s^\rho\ K(s)\ ds}{\rho!}\ g^{(\rho)}(x)\ h^\rho,$$

uniformly on any compact interval.

In the following, a ρ-regularity assumption for g (and also f) will always be associated with the use of such a kernel K of order ρ.

3.2 Interval estimation with local bootstrap

Interval estimation of $r(x) = \mathbb{E}(Y_0|X_0 = x)$ has been proposed with local bootstrap schemes, see [16], [19] and also [17]. All these cited references assume an α- or β-mixing condition for the stationary underlying process.

The local bootstrap for nonparametric regression is defined as follows. The given data is $Z_1 = (X_1, Y_1), \ldots, Z_n = (X_n, Y_n)$ with $Y_t \in \mathbb{R}$, $X_t \in \mathbb{R}$ $(t = 1, \ldots, n)$. As mentioned at the beginning of section 3, the results easily extend to the case where $X_t \in \mathbb{R}^d$ with $1 \le d < \infty$. Consider the empirical distribution function for Y_t given $X_t = x$,

$$\hat{F}_b(.|x) = \sum_{t=1}^{n} K(\frac{x - X_t}{b}) \mathbb{1}_{\{Y_t \le .\}} / \hat{f}_{n,b}(x),$$

with kernel $K(.)$ and estimator $\hat{f}_{n,b}(.)$ as in section 3.1, but with bandwidth b generally different from h used in the estimator $\hat{r}_{n,h}$. Construct the bootstrap sample as

$$(X_1, Y_1^*), \ldots, (X_n, Y_n^*),\ Y_t^* \sim \hat{F}_b(.|X_t)\ (t = 1, \ldots, n), \tag{3.1}$$

such that given Y_t^* is independent of Y_s^* $(s \ne t)$, given $X_1, \ldots, X_n$. Note that the bootstrap sample in (3.1) involves only resampling of the response variables Y_t^* $(t = 1, \ldots, n)$ in an independent way. For the particular problem of bootstrapping nonparametric estimates of conditional expectations, this turns out to be sufficient, although the underlying process can be very general. The reason for this is the 'whitening by windowing' principle mentioned already in section 1: as will be shown in Theorem 4.3 and Proposition 5.2, the asymptotic distribution of $\hat{r}(x)$ is the same as in the independent case, thus indicating that a bootstrap mechanism does not need to mimic dependence properties. In the sequel, bootstrap moments and distributions induced by the resampling random mechanism in (3.1) are always equipped with an asterisk *.

By the plug-in principle, we consider the bootstrapped estimator

$$\hat{r}_{n,h}^*(x) = \frac{\hat{g}_{n,h}^*(x)}{\hat{f}_{n,h}(x)},\ \hat{g}_{n,h}^*(x) = \frac{1}{nh}\sum_{t=1}^{n} K(\frac{x - X_t}{h}) Y_t^*.$$

Bootstrap percentile confidence intervals for $\mathbb{E}[\hat{r}(x)]$ or $r(x)$ are then constructed as usual.

3.2.1 Numerical results

We briefly investigate the finite sample behavior of the local bootstrap; asymptotic properties are given in section 4.5. Consider the following bilinear process

$$X_t = 0.5\xi_{t-1}X_{t-1} + \xi_t, \tag{3.2}$$

where $(\xi_t)_{t\in\mathbb{Z}}$ is an i.i.d. innovation sequence with ξ_t independent from $\{X_s; s < t\}$. We consider two models,

(M1): process in (3.2) with ξ_t i.i.d. $\sim$ Rademacher, i.e., $\mathbb{P}[\xi_t = 1] = \mathbb{P}[\xi_t = -1] = 1/2$.

(M2): process in (3.2) with ξ_t i.i.d. $\sim$ Uniform([-1,1]).

Figure 3.1 shows a realization of each model with sample size $n = 512$, including the lag(1) scatter-plot of X_t versus X_{t-1} and the estimator $\hat{r}_{n,h}(.)$ using the standard-Gaussian kernel for bandwidth $h = 0.25$. The qualitative difference of the models is the discrete versus the continuous innovations in

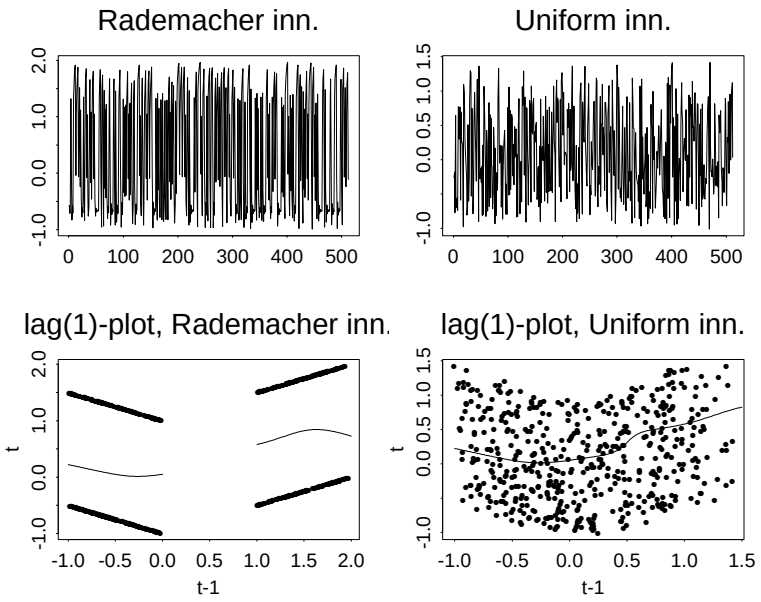

Figure 3.1: Top row: One realization for each model (M1) and (M2) with $n = 512$. Bottom row: For each model (M1) and (M2), lag(1) scatter-plot of X_t versus X_{t-1} with the line indicating the curve estimator $\hat{r}_{n,h}(.)$ with the standard-Gaussian kernel for $h = 0.25$.

(M1) and (M2), respectively. Also, (M2) is β-mixing, whereas (M1) is not even α-mixing.

We consider the local bootstrapped estimator $\hat{r}^*_{n,h}(x)$ with two pilot-bandwidths $b = 0.25$ and $b = 0.5$ in (3.1); we don't vary here over the bandwidth h and assume it to be reasonable and pre-given from the application, see also bottom row of Figure 3.1. For (M1), we consider $x = 1.60$ and for (M2) $x = 0.76$: both correspond approximately to the 80% quantile of the marginal distribution of the underlying process. We always draw 100 bootstrap replicates and the simulation experiments run over 50 independent realizations of the model specifications in (M1) and (M2), respectively.

Figure 3.2 shows the quality of the bootstrap approximation for the distribution of $\hat{r}_{n,h}(x)$. We are here primarily interested in the goodness of the approximation of the centered estimator, thus

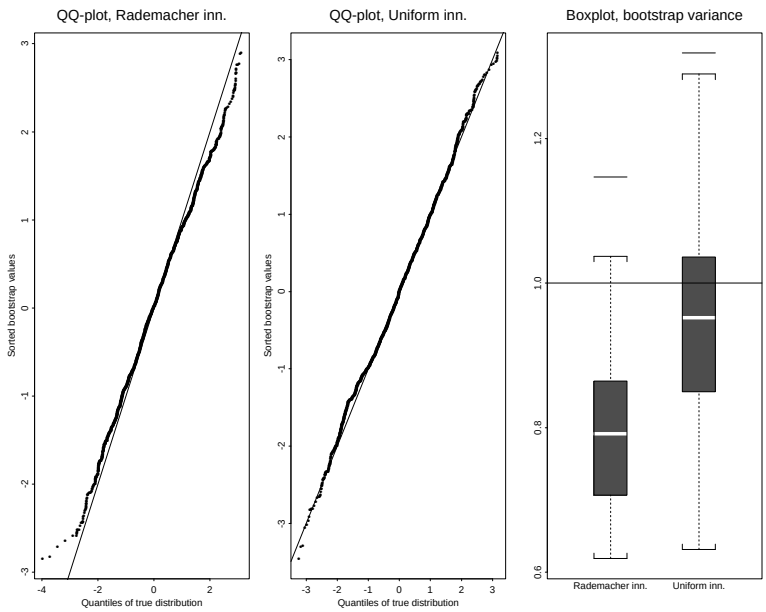

Figure 3.2: Two left panels: QQ-plot for the bootstrap values in (3.3) compared to the true distribution of the random variable in (3.4); the target is indicated by the line. Right panel: Boxplot of 50 bootstrap variance estimates based on (3.3) for the true variance in (3.4), which is indicated by the horizontal line. Everything with $n = 512, h = 0.25, b = 0.5$.

ignoring bias issues. We look at

$$\left(\hat{r}^*_{n,h}(x) - \mathbb{E}^*(\hat{r}^*_{n,h}(x))\right)/\sigma(x),\ \ \sigma^2(x) = \mathrm{Var}(\hat{r}_{n,h}(x)) \tag{3.3}$$

as an approximation of

$$\left(\hat{r}_{n,h}(x) - \mathbb{E}(\hat{r}_{n,h}(x))\right)/\sigma(x). \tag{3.4}$$

The standardization with $\sigma(x)$ is used so that the quantities are on the same scale for both models (M1) and (M2). The two left hand panels in Figure 3.2 show for each model (M1) and (M2) the QQ-plots for the bootstrap values of the quantities in (3.3) (100 bootstrap replicates for each of the 50 model simulations) compared to the true distributions appearing in (3.4). The right hand panel shows the 50 bootstrap variance estimates (for the 50 simulations) based on the quantities in (3.3) for the true variance in (3.4) which is 1.

Table 3.1 reports the bootstrap variance estimates based on (3.3) for the true variance 1 in (3.4). Denote by

$$\hat{\sigma}^2_{stand} = \mathrm{Var}^*(\hat{r}^*_{n,h}(x))/\sigma^2(x),$$

and by $MSE(\hat{\sigma}^2_{stand}) = \mathbb{E}((\hat{\sigma}^2_{stand} - 1)^2)$.

On the relative scale, when standardized by $\sigma(x)$ in (3.3) and (3.4), the accuracy for the non-mixing model with Rademacher innovations (M1) is worse than for the mixing model with continuous innovations (M2). In absolute terms, the model (M1) has with $\sigma(1.60) = 0.0064$ a much larger variability than (M2) with $\sigma(0.76) = 0.0019$ and so do also the bootstrap estimates. In this small

model and bandwidths	bias $= \mathbb{E}(\hat{\sigma}^2_{stand}) - 1$	$(\mathrm{Var}(\hat{\sigma}^2_{stand}))^{1/2}$	$MSE(\hat{\sigma}^2_{stand})$
(M1); h=0.25,b=0.25	-0.21	0.11	0.06 (0.006)
(M1); h=0.25,b=0.5	0.20	0.11	0.05 (0.006)
(M2); h=0.25,b=0.25	-0.07	0.16	0.03 (0.005)
(M2); h=0.25,b=0.5	0.04	0.16	0.03 (0.005)

Table 3.1: Local bootstrap variance estimates, sample size $n = 512$.

experiment, no systematic behavior of the effect of the pilot-bandwidth b is found: when testing the null-hypothesis that the MSE is the same for $b = 0.25$ and $b = 0.5$, the paired two-sample Wilcoxon-test yields the non-significant p-values 0.07 in both models (M1) and (M2). The relative mean squared error for the bootstrap variance estimators is between 3% and 6%, indicating a good performance, see Table 3.1.

4 Asymptotic properties

For asymptotic analysis we assume that

$$\begin{aligned}
&\text{the marginal density } f(.) \text{ of } X_t \text{ exists and is continuous,}\\
&f(x) > 0 \text{ for the point } x \text{ of interest,}\\
&\text{the function } r(.) = \mathbb{E}(Y_0|X_0 = .) \text{ exists and is continuous,}\\
&\text{for some } p \geq 1,\ g_p(.) = f(.)\mathbb{E}(|Y_0|^p|X_0 = .) \text{ exists and is continuous.}
\end{aligned} \tag{4.1}$$

We set $g = fr$ with obvious short notation. Moreover, assume one of the following moment assumptions. Either,

$$\mathbb{E}(|Y_0|^S) < \infty, \text{ for some } S \geq p \text{ with } p \text{ as in (4.1)}, \tag{4.2}$$

or

$$\mathbb{E}(\exp(|Y_0|)) < \infty. \tag{4.3}$$

We also consider the following conditionally centered equivalent of g_2 appearing in the asymptotic variance of the estimator $\hat{r}$,

$$G_2(x) = f(x)\mathrm{Var}\,(Y_0|X_0 = x) = g_2(x) - f(x)r^2(x).$$

4.1 Variance and asymptotic normality of $\hat{g}$

We consider first the properties of $\hat{g}(x))$. Notice that $g = f$ if $Y_0 \equiv 1$.

Denote by $f_{(k)}$ the density of the pair (X_0, X_k) and assume that there exists some constant $C > 0$ such that,

$$\begin{aligned}
&\sup_{k\in\mathbb{N}} \|f_{(k)}\|_\infty \leq C, \text{ and}\\
&r_{(k)}(x, x') = \mathbb{E}\left(|Y_0Y_k|\big|X_0 = x, X_k = x'\right) \text{ are continuous, uniformly over all } k \in \mathbb{N}.
\end{aligned} \tag{4.4}$$

Under this assumption, the functions $g_{(k)} = f_{(k)}r_{(k)}$ are locally bounded. The following result extends Lemma 1 in [8], valid for the case of $\hat{f}$, to the estimate $\hat{g}$ under weak dependence with either ψ_1 or ψ_2. Also note that replacing the sequence Y_t by $Y_t - r(x)$ at some fixed point x yields asymptotics for $\hat{g}(x) - r(x)\hat{f}(x)$ which will be useful for the analysis of $\hat{r}(x)$.

Proposition 4.1 *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying the conditions (4.1) with $p = 2$, (4.3) and (4.4). Suppose $n^\delta h \to \infty$ for some $\delta \in]0,1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > 2 + j$, for $j = 1$ or $j = 2$. Then, uniformly in x belonging to any compact subset of $\mathbb{R}$,*

$$\mathrm{Var}(\hat{g}(x)) = \frac{1}{nh}\ g_2(x) \int K^2(u)du\ +\ o\left(\frac{1}{nh}\right)$$

and

$$\mathrm{Var}\left(\hat{g}(x) - r(x)\hat{f}(x)\right) = \frac{1}{nh} G_2(x) \int K^2(u)du + o\left(\frac{1}{nh}\right).$$

Remark 4.1. The moment condition in (4.3) can be relaxed. Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the condition (4.1) with $p = 2$, (4.2) with $S > 2$ and (4.4). Let $n^\delta h \to \infty$ for some $\delta \in]0,1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > \frac{(2+j)S-4}{S-2} + \frac{2}{\delta(S-2)}$, for $j = 1$ or $j = 2$. Then, the assertions of Proposition 4.1 hold.

We now turn to central limit theorem questions. As mentioned in section 1, ψ'-weak dependence conditions yield more attractive and simple results. Nevertheless, the more standard ψ_1 and ψ_2 weak dependence conditions also lead to CLT's, see Proposition 5.1 in section 5.1.

Theorem 4.1 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the condition (4.1) with $p = 2$, (4.3) and (4.4). Suppose that $n^\delta h \to \infty$ for some $\delta \in]0,1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi'_j, \mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > 2 + j$, for $j = 1$ or $j = 2$. Then*

$$\sqrt{nh}\ \ (\hat{g}(x) - \mathbb{E}\hat{g}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, g_2(x) \int K^2(u)du\right)$$

and

$$\sqrt{nh}\ \ \left([\hat{g}(x) - r(x)\hat{f}(x)] - \mathbb{E}[\hat{g}(x) - r(x)\hat{f}(x)]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, G_2(x)\ \int K^2(u)du\right).$$

Remark 4.2. The results stated in Theorem 4.1 also hold for finite dimensional convergence; the components are asymptotically mutually independent, as for i.i.d samples.

In order to consider asymptotics for the ratio estimator $\hat{r}$ one needs to study higher order asymptotics, using ideas from [3], which is the topic of the next section.

4.2 Higher order moments of $\hat{g}$

We give here rate-optimal bounds for higher order moments which turn out to be useful for asymptotics of $\hat{r}$, deriving an exponential inequality and almost sure convergence results.

Theorem 4.2 *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying the conditions (4.1) with $p = 2$, (4.3) and (4.4). Let $q \geq 3$ be some integer. Suppose that $n^\delta h \to \infty$ for some $\delta \in]0,1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi'_j, \mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > (2 + j)(q - 1)$, for $j = 1$ or $j = 2$. Then, for all x belonging to some compact set,*

$$\overline{\lim}_{n\to\infty} (nh)^{q/2} \left|\mathbb{E}\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)^q\right| < \infty.$$

Remark 4.3. For ψ-weak dependence we obtain the following. Assume the conditions of Theorem 4.2 but replace ψ'-weak dependence by one of the following:

(i) The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_1,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and

$$a > a_1(q,\delta) = \max\left(q-1, \frac{6\delta(q-1)}{(q-2)+\delta(4-q)}\right).$$

(ii) The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_2,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and

$$a > a_2(q,\delta) = \max\left(q-1, \frac{8\delta(q-1)}{(q-2)+\delta(4-q)}\right).$$

Then, the assertion of Theorem 4.2 holds true.

4.3 Asymptotic normality of the regression

Proposition 4.2 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the conditions (4.1) with $p=2$, (4.3) and (4.4). Consider a positive kernel K. Let $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0,2]$, and $n^\delta h \to \infty$ for some $\delta \in]0,1[$. In addition, assume one of the following.*

(i) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_1,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > \max(3, 9\delta)$.*

(ii) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_2,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > \max(3, 12\delta)$.*

Then, uniformly in x belonging to any compact subset of $\mathbb{R}$,

$$\mathbb{E}\Big(\hat{r}(x)\Big) = r(x) + O\left(h^\rho + \frac{1}{nh}\right).$$

Proposition 4.3 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the conditions (4.1) with $p=2$, (4.3) and (4.4). Consider a positive kernel K. Let $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0,2]$, and $n^\delta h \to \infty$ for some $\delta \in]0,1[$. In addition, assume one of the following.*

(i) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_1,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > \max\left(5, \frac{30\delta}{7-5\delta}\right)$.*

(ii) *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_2,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > \max\left(5, \frac{40\delta}{7-5\delta}\right)$.*

Then, uniformly in x belonging to any compact subset of $\mathbb{R}$,

$$\mathrm{Var}\Big(\hat{r}(x)\Big) = \frac{G_2(x)}{nhf^2(x)}\int K^2(u)du + o\left(\frac{1}{nh}\right).$$

Theorem 4.3 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the conditions (4.1) with $p=2$, (4.3) and (4.4). Consider a positive kernel K. Let $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0,2]$, and $nh^{1+2\rho} \to 0$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_j,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and*

$$a > \min\left(\max\left(2+j, 3(2+j)\delta\right), \max\left(2+j+\frac{1}{\delta}, \frac{2+2(2+j)\delta}{1+\delta}\right)\right),$$

for $j=1$ or $j=2$. Then, for all x belonging to any compact subset of $\mathbb{R}$,

$$\sqrt{nh}\,\Big(\hat{r}(x) - r(x)\Big) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \frac{G_2(x)}{f^2(x)}\int K^2(u)du\right).$$

In case where the bandwidth satisfies only $n^\delta h \to \infty$ for some $\delta \in]0,1[$, asymptotic normality when centered at $\mathbb{E}[\hat{r}(x)]$ still holds, see Proposition 5.2 in section 5.3.

4.4 Almost sure convergence properties

We now give an almost sure convergence result which can be deduced from Theorem 4.2.

Theorem 4.4 *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying the conditions (4.1) with $p=2$, (4.3), (4.4) and that it is either $(\theta,\psi_1,\mathcal{L})$- or $(\theta,\psi_2,\mathcal{L})$-weak dependent with $\theta_r \le a^r$ for some $0<a<1$.*

(i) *If $nh/\log^4(n)\to\infty$, then for any $M>0$, almost surely,*

$$\sup_{|x|\le M}|\hat{g}(x)-\mathbb{E}\hat{g}(x)| = O\left(\frac{\log^2(n)}{\sqrt{nh}}\right).$$

(ii) *For any $M>0$, if $f,g\in\mathcal{C}_\rho$ for some $\rho\in]0,\infty[$, $h\sim\left(\frac{\log^4(n)}{n}\right)^{1/(1+2\rho)}$ and $\inf_{|x|\le M} f(x)>0$, then, almost surely,*

$$\sup_{|x|\le M}|\hat{r}(x)-r(x)| = O\left\{\left(\frac{\log^4(n)}{n}\right)^{\rho/(1+2\rho)}\right\}.$$

Remark 4.4. Under the conditions of Theorem 4.4 **(ii)**, but assuming only the weaker condition $n^\delta h\to\infty$ for some $\delta\in]0,1[$, we obtain

$$\sup_{|x|\le M}|\hat{r}(x)-r(x)| = o(1) \text{ almost surely.}$$

4.5 Validity of the local bootstrap

For the asymptotic validity of the local bootstrap we make the following assumptions. Consider the conditional variance and centered 4th-order moment,

$$V(x) = \mathrm{Var}(Y_0|X_0=x),$$
$$M(x) = \mathbb{E}\left((Y_0-\mathbb{E}(Y_0|X_0=x))^4|X_0=x\right)$$

We assume that in a neighborhood U_x of x,

$$V(.) \text{ is differentiable with } \sup_{v\in U_x}|V'(v)|<\infty, \tag{4.5}$$
$$M(.) \text{ is differentiable with } \sup_{v\in U_x}|M'(v)|<\infty. \tag{4.6}$$

Moreover, denote by

$$Z_t^{(s,1)} = (Y_t^s, X_t),\ s\in\mathbb{N}.$$

Theorem 4.5 *Suppose that $(Z_t)_{t\in\mathbb{Z}}$ satisfies the assumptions of Proposition 5.2 in section 5.3, and the stationary sequences $(Z_t^{(s,1)})_{t\in\mathbb{Z}}$ satisfy the weak dependence assumptions of Remark 4.4 for $s=1,2,3,4$. Assume additionally that (4.5) and (4.6) hold, $n^\delta h\to\infty$ for some $\delta\in]0,1[$, and the pilot bandwidth is $b=O(n^{-\beta})$ for some $\beta\in]0,1/3[$. Then,*

(i)

$$\sup_{v\in\mathbb{R}}\left|\mathbb{P}^*\left(\sqrt{nh}(\hat{r}^*_{n,h}(x)-\mathbb{E}^*(\hat{r}^*_{n,h}(x)))\le v\right)-\mathbb{P}\left(\sqrt{nh}(\hat{r}_{n,h}(x)-\mathbb{E}(\hat{r}_{n,h}(x)))\le v\right)\right| = o_P(1).$$

(ii) *In case of no asymptotic bias* $\lim \sqrt{nh}\,(\mathbb{E}(\hat{r}_{n,h}(x)) - r(x)) = 0$, *i.e.* $nh^{1+2\rho} \to 0$ *if* $f, g \in \mathcal{C}_\rho$ *for some* $\rho \in]0, \infty[$, *the* $\mathbb{E}(\hat{r}_{n,h}(x))$ *term in* **(i)** *can be replaced by* $r(x)$.

Remark 4.5. In the case of asymptotic bias in **(ii)**, the local bootstrap also works if additional conditions on the kernel K and on the regularity of the function r hold, and if the pilot bandwidth b satisfies $b/h \to \infty$ $(n \to \infty)$, see for example [19]. Then

$$\sup_{v \in \mathbb{R}} \left| \mathbb{P}^* \left(\sqrt{nh}(\hat{r}^*_{n,h}(x) - \hat{r}_{n,b}(x)) \le v \right) - \mathbb{P} \left(\sqrt{nh}(\hat{r}_{n,h}(x) - r(x)) \le v \right) \right| = o_P(1).$$

Remark 4.6. Weak dependence of $(Z_t^{(s,1)})$ $(s = 1, 2, 3, 4)$ holds if $(Y_t)_{t \in \mathbb{Z}}$ is a suitably regular Bernoulli shift and $X_t = Y_{t-\ell}$ for some $\ell \in \mathbb{N}$. This, because for $Y_t = F(\xi_{t-j}, j \in \mathbb{Z})$

$$\begin{aligned} & |F^2(\xi_{t-j}, j \in \mathbb{Z}) - F^2(\xi_{t-j}\mathbb{I}_{|j|<k}, j \in \mathbb{Z})| \\ = \; & |F(\xi_{t-j}, j \in \mathbb{Z}) - F(\xi_{t-j}\mathbb{I}_{|j|<k}, j \in \mathbb{Z})|\, |F(\xi_{t-j}, j \in \mathbb{Z}) + F(\xi_{t-j}\mathbb{I}_{|j|<k}, j \in \mathbb{Z})|. \end{aligned}$$

Assuming some regularity similar as in the discussion following (2.2) and $\mathbb{E}|Y_t|^{1+\kappa} < \infty$ for some $0 < \kappa < \infty$ yields by Hölder's inequality weak dependence for (Y_t^2), typically with a slower decay of θ_r. By finite iteration, weak dependence of (Y_t^s) follows for $s \in \mathbb{N}$.

5 Proofs

5.1 Variance and asymptotic normality of $\hat{g}$

Proof of Proposition 4.1. Denote by C (different) constants whose values are allowed to change. Let

$$T_t(x) = Y_t\ \mathbb{I}_{\{|Y_t| \le M(n)\}} K\left(\frac{x - X_t}{h}\right)\ (0 \le t \le n). \tag{5.1}$$

Then the truncated kernel estimator of $g(x)$,

$$\tilde{g}(x) = \frac{1}{nh} \sum_{t=1}^{n}\ T_t(x), \tag{5.2}$$

satisfies

$$\begin{aligned} (nh)\mathbb{E}\Big(\hat{g} - \tilde{g} - \mathbb{E}\,(\hat{g} - \tilde{g})\Big)^2(x) \;&\le\; \frac{2n^2}{nh}\,\mathbb{E}\left[Y_0^2 \mathbb{I}_{\{|Y_0| > M(n)\}} K^2\left(\frac{x - X_0}{h}\right)\right] \\ &\le\; \mathrm{C}\ M^{(2-S)}(n) nh^{-1}\ \to 0\ (n \to \infty) \end{aligned}$$

for $M(n) = M_0 n^\gamma, M_0 > 0, \gamma > (1 + \delta)/(S - 2)$. It remains to estimate

$$\mathrm{Var}(\tilde{g}(x)) = \frac{1}{nh^2}\mathrm{Var}T_0(x) + \frac{2}{n^2h^2}\sum_{r=1}^{n-1}(n - r)\mathrm{Cov}\,(T_0(x), T_r(x))\,.$$

A classical result (see for example [21], p. 37) shows that

$$\mathrm{Var}(T_0(x)) = hg_2(x)\int K^2(u)\ du\ +\ o(h).$$

It follows from the boundedness assumptions on densities that

$$|\mathrm{Cov}\,(T_0(x), T_r(x))\,| \le \mathrm{C}h^2.$$

Moreover, the $(\theta, \psi_1, \mathcal{L})$-weak dependence assumption yields,

$$|\mathrm{Cov}\,(T_0(x), T_r(x))\,| \le \mathrm{C}\frac{\theta_r M^2(n)}{h}. \tag{5.3}$$

Next, we use a truncation device due to Tran [28]: if $a > \frac{3S-4}{S-2} + \frac{2}{\delta(S-2)}$, there exists $\varsigma \in]0,1[$ such that $(2\delta + 2\gamma)/(a-1) < \varsigma < \delta$, so that

$$\frac{2(nh)}{n^2h^2}\sum_{r=1}^{[n^\varsigma]}(n-r)|\mathrm{Cov}\,(T_0(x), T_r(x))\,| \le \mathrm{C}\ n^{\varsigma-\delta} \to 0\ (n\to\infty)$$

and

$$\begin{aligned}\frac{2(nh)}{n^2h^2}\sum_{r=1+[n^\varsigma]}^{n-1}(n-r)|\mathrm{Cov}\,(T_0(x), T_r(x))\,| &\le \mathrm{C}\ \frac{M^2(n)}{h^2}\sum_{r\ge[n^\varsigma]}\theta_r \\ &\le \mathrm{C}\ n^{\varsigma(1-a)+2\delta-2\gamma} \to 0\ (n\to\infty).\end{aligned}$$

Using the bounds given above, the first assertion of Proposition 4.1 follows. For a $(\theta, \psi_2, \mathcal{L})$-weak dependent process, the result follows from

$$|\mathrm{Cov}\,(T_0(x), T_r(x))\,| \le \mathrm{C}\ \frac{\theta_r M^2(n)}{h^2}.$$

in place of (5.3).

The second assertion follows by replacing Y_t with $Y_t - r(x)$. □

Proof of Theorem 4.1. We proceed as in [22] and more specifically as in [5] for density estimation.

Consider a sequence $(W_n)_{n\in\mathbb{N}}$ of i.i.d. r.v.'s with standard normal law, and independent from $((X_t, Y_t))_{t\in\mathbb{Z}}$. Set $M(n) = \log(n)$, $nh\sigma_n^2 = \mathrm{Var}\,(\sum_{t=1}^n T_t(x))$ with $T_t(x)$ given in (5.1), and define

$$\begin{aligned}\xi_t &= \frac{1}{\sigma_n\sqrt{nh}}\left(T_t(x) - \mathbb{E}T_t(x)\right), \\ S_k &= \sum_{t=1}^{k}\xi_t,\ 1\le k\le n,\ \text{and } S_0 = 0, \\ \tau_k &= \sum_{t=k}^{n}V_t,\ 1\le k\le n,\ \text{and } \tau_{n+1} = 0,\end{aligned}$$

where $\frac{v_k}{\sigma_n^2} = |\mathrm{Var}(S_k) - \mathrm{Var}(S_{k-1})|$ and $V_k = \frac{\sqrt{v_k}}{\sigma_n}W_k$. For the Lindeberg method let φ denote some three times differentiable function with bounded derivatives up to order 3, and consider the following:

$$U_t = S_{t-1} + \tau_{t+1},\ R_t(x) = \varphi\,(U_t + x) - \varphi\,(U_t) - \frac{v_t}{2\sigma_n^2}\varphi''\,(U_t)\ \ (1\le t\le n).$$

Now, we want to establish that

$$\sigma_n \to \left(g_2(x) \int K^2(u)du \right)^{1/2}, \tag{5.4}$$

$$\sqrt{nh}\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right) - \sigma_n S_n \overset{\mathbb{L}^2}{\to} 0, \tag{5.5}$$

$$S_n \overset{\mathcal{D}}{\to} \mathcal{N}(0,1). \tag{5.6}$$

We consider either a $(\theta, \psi_1, \mathcal{L})$- or $(\theta, \psi_1', \mathcal{L})$-weak dependent sequence $(Z_t)_{t\in\mathbb{Z}}$ and stress on their different bounds. Analogously as in the proof of Proposition 4.1, we show that formula (5.4) holds true if the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_1', \mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $M^2(n)h_n^{(a-3)} \to 0$. The same result is valid in the $(\theta, \psi_1, \mathcal{L})$-weak dependent case.

Formula (5.5) is easily proved using the exponential moment assumption (4.3).

In order to prove formula (5.6), we apply the so-called Lindeberg-Rio method [23]. Clearly,

$$\left|\mathbb{E}\varphi\left(S_n\right) - \mathbb{E}\varphi\left(W_0\right)\right| \leq \sum_{t=1}^{n} \left|\mathbb{E}R_t(\xi_t)\right| + \sum_{t=1}^{n} \left|\mathbb{E}R_t(V_t)\right|. \tag{5.7}$$

Since

$$\begin{aligned}
\left|\mathbb{E}R_t(V_t)\right| &= \left|\mathbb{E}\left(\varphi\left(U_t+V_t\right) - \varphi\left(U_t\right) - V_t\varphi'\left(U_t\right) - \frac{V_t^2}{2}\varphi''\left(U_t\right)\right)\right| \\
&\leq \frac{1}{6}\mathbb{E}\left|V_t^3\varphi^{(3)}\left(U_t+\vartheta V_t\right)\right|, \text{ with } 0<\vartheta<1, \\
&\leq \frac{\|\varphi^{(3)}\|_\infty}{6}\left(\frac{v_t}{\sigma_n^2}\right)^{3/2}\mathbb{E}\left|W_0\right|^3,
\end{aligned}$$

and

$$\begin{aligned}
\frac{v_t}{\sigma_n^2} &= \left|\mathrm{Var}\left(S_t\right) - \mathrm{Var}\left(S_{t-1}\right)\right| \\
&\leq C\left(\frac{1}{n} + \sum_{j=1}^{t-1}\frac{1}{nh}\min\left(\frac{M^2(n)\theta_{t-j}}{h}, h^2\right)\right),
\end{aligned}$$

we obtain

$$\sum_{t=1}^{n}\left|\mathbb{E}R_t(V_t)\right| \leq \frac{C}{\sigma_n^3}\sum_{t=1}^{n} v_t^{3/2} \leq \frac{C}{n^{3/2}\sigma_n^3}\sum_{t=1}^{n}\left(1 + \sum_{j=1}^{t-1}\min\left(\frac{M^2(n)\theta_j}{h^2}, h\right)\right)^{3/2}. \tag{5.8}$$

Moreover,

$$\begin{aligned}
R_t(\xi_t) &= \varphi\left(U_t+\xi_t\right) - \varphi\left(U_t\right) - \frac{v_t}{2\sigma_n^2}\varphi''\left(U_t\right) \\
&= \xi_t\varphi'\left(U_t\right) + \frac{1}{2}\left(\xi_t^2 - \frac{v_t}{2\sigma_n^2}\right)\varphi''\left(U_t\right) + \frac{1}{6}\xi_t^3\varphi^{(3)}\left(U_t+\vartheta_t\xi_t\right), \text{ with } 0<\vartheta_t<1.
\end{aligned}$$

It then follows that

$$\sum_{t=1}^{n}\left|\mathbb{E}R_t(\xi_t)\right| \leq \left|\mathbb{E}\left(\sum_{t=1}^{n}\mathrm{Cov}\left(\xi_0,\xi_t\right)\sum_{k=t+1}^{n}\varphi''\left(U_t\right)\right)\right| + \frac{1}{2}\left|\sum_{t=1}^{n}\mathbb{E}\left(\varphi''\left(U_t\right)\left(\xi_t^2 - \mathbb{E}\xi_t^2\right)\right)\right|$$

$$
\begin{aligned}
&+ \left| \sum_{t=1}^{n} \sum_{j=1}^{t-1} \mathrm{Cov}\left(\varphi''\left(S_{t-1-j}+\tau_{j+1}\right)\xi_{t-j},\xi_t\right)\right| \\
&+ \frac{1}{2}\left| \sum_{t=1}^{n} \sum_{j=1}^{t-1} \mathrm{Cov}\left(\varphi^{(3)}\left(S_{t-1-j}+\tau_{j+1}+\vartheta_{t-j}\xi_{t-j}\right)\xi_{t-j}^2,\xi_t\right)\right| \\
&+ \frac{1}{6}\left| \sum_{t=1}^{n} \mathbb{E}\left(\varphi^{(3)}\left(U_t+\iota_t\xi_t\right)\xi_t^3\right)\right| = E_1+E_2+E_3+E_4+E_5. \qquad (5.9)
\end{aligned}
$$

We now follow [5] to bound the five terms above.

$$
E_1 \le \sum_{t=1}^{n} \left|\mathrm{Cov}\left(\xi_0,\xi_t\right)\right| \left| \sum_{k=t+1}^{n} \mathbb{E}\varphi''\left(U_t\right)\right| \le \frac{C}{nh\sigma_n^2} \sum_{t=1}^{n} (n-t)\min\left(\frac{M^2(n)\theta_t}{h}, h^2\right) \qquad (5.10)
$$

$$
\begin{aligned}
E_2 &\le \frac{1}{2}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\left|\mathrm{Cov}\left(\varphi^{(3)}\left(S_{j-1}+\iota_j\xi_j+\tau_{k+1}\right)\xi_j,\frac{\xi_t^2}{2}\right)\right| + \sum_{t=1}^{n}\left|\mathrm{Cov}\left(\varphi''(\tau_{k+1}),\frac{\xi_t^2}{2}\right)\right| \\
&\le \begin{cases} \frac{CM(n)}{\left(\sigma_n\sqrt{nh}\right)^3}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\min\left(\frac{(t-j)M^2(n)\theta_j}{h},h^2\right)+\frac{Ch}{\sigma_n^2}, & \text{in the } (\theta,\psi_1,\mathcal{L}) \text{ case} \\ \frac{CM(n)}{\left(\sigma_n\sqrt{nh}\right)^3}\sum_{t=2}^{n}\sum_{j=1}^{t-1}\min\left(\frac{M^2(n)\theta_j}{h},h^2\right)+\frac{Ch}{\sigma_n^2}, & \text{in the } (\theta,\psi_1',\mathcal{L}) \text{ case} \end{cases} \qquad (5.11)
\end{aligned}
$$

$$
E_3 \le \begin{cases} \sum_{t=1}^{n}\sum_{j=1}^{t-1}\frac{C}{nh\sigma_n^2}\min\left(\frac{M^2(n)(t-j)\theta_j}{h},h^2\right), & \text{in the } (\theta,\psi_1,\mathcal{L}) \text{ case} \\ \sum_{t=1}^{n}\sum_{j=1}^{t-1}\frac{C}{nh\sigma_n^2}\min\left(\frac{M^2(n)\theta_j}{h},h^2\right), & \text{in the } (\theta,\psi_1',\mathcal{L}) \text{ case} \end{cases} \qquad (5.12)
$$

$$
E_4 \le \begin{cases} \sum_{t=1}^{n}\sum_{j=1}^{t-1}\frac{C}{\left(\sigma_n\sqrt{nh}\right)^3}\min\left(\frac{M^3(n)(t-j)\theta_j}{h},h^2M(n)\right), & \text{in the } (\theta,\psi_1,\mathcal{L}) \text{ case} \\ \sum_{t=1}^{n}\sum_{j=1}^{t-1}\frac{C}{\left(\sigma_n\sqrt{nh}\right)^3}\min\left(\frac{M^3(n)\theta_j}{h},h^2M(n)\right), & \text{in the } (\theta,\psi_1',\mathcal{L}) \text{ case} \end{cases} \qquad (5.13)
$$

$$
E_5 \le \frac{C}{(\sigma_n\sqrt{nh})^3}\sum_{t=1}^{n} M(n)h \le \frac{CM(n)}{\sigma_n^3\sqrt{nh}}. \qquad (5.14)
$$

If the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_1',\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ for some $a3$, then by using (5.8)-(5.14), the right hand side of (5.7) tends to zero with n^{-1}.

For a $(\theta,\psi_1,\mathcal{L})$-weak dependent sequence, using again (5.8)-(5.14), we need $\theta_r = O(r^{-a})$ with

$$
a > \max\left(3+\frac{1}{\delta}, \frac{6\delta+2}{\delta+1}\right).
$$

All together, this completes the proof of the first assertion. The second assertion is an application of the first one, where Y_t is replaced with $Y_t - r(x)$. □

Remark 5.1. The CLT can also be proved by Bernstein's blocking technique, as in [8]. Then both assertions in Theorem 4.1 hold true in each of the following cases.

(i) The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_1,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$, where $a > \max\Big(3,\ 9\delta\Big)$.

(ii) The sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_2,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$, where $a > \max\Big(4,\ 12\delta\Big)$.

The foregoing allows to state another result, given in Proposition 5.1 below. Define

$$\alpha_1(\delta) = \min\left(\max(3,9\delta), \max\left(3+\frac{1}{\delta}, \frac{6\delta+2}{\delta+1}\right)\right) = \begin{cases} 3 & \text{if } 0<\delta<\frac{1}{3}, \\ 9\delta & \text{if } \frac{1}{3}\le\delta<\frac{1+\sqrt{5}}{6}, \\ 3+\frac{1}{\delta} & \text{otherwise.}\end{cases}$$

and

$$\alpha_2(\delta) = \min\left(\max(4,12\delta), \max\left(4+\frac{1}{\delta}, \frac{8\delta+2}{\delta+1}\right)\right) = \begin{cases} 4 & \text{if } 0<\delta<\frac{1}{3}, \\ 12\delta & \text{if } \frac{1}{3}\le\delta<\frac{1}{2}, \\ 4+\frac{1}{\delta} & \text{otherwise.}\end{cases}$$

Note that $3 \le \alpha_1(\delta) \le \frac{3(1+\sqrt{3})}{2}$ and $4 \le \alpha_2(\delta) \le 6$.

Proposition 5.1 *Assume the conditions of Theorem 4.1, but require instead that $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_j,\mathcal{L})$-weak dependent with $\theta_r = O(r^{-a})$ and $a > \alpha_j(\delta)$, for $j=1$ or $j=2$. Then, both assertions* **(i)** *and* **(ii)** *of Theorem 4.1 hold.*

A proof of Proposition 5.1 is omitted and we refer to [8] for the essential ideas.

5.2 Higher order moments of $\hat{g}$

Proof of Theorem 4.2. We keep the notations from (5.1) and (5.2) and denote again by C a universal constant (whose value might change). Since $\mathbb{E}(\exp(|Y_0|)) < \infty$,

$$\begin{aligned}(nh)^{q/2}\mathbb{E}|\hat{g}-\tilde{g}-\mathbb{E}(\hat{g}-\tilde{g})|^q(x) &\le \frac{(2n)^q}{(nh)^{q/2}}\,\mathbb{E}\left[|Y_0|^q \mathbb{I}_{\{|Y_0|>M(n)\}}\left|K\left(\frac{x-X_0}{h}\right)\right|^q\right] \\ &\le \mathrm{C}\frac{(2n)^q}{(nh)^{q/2}}\|u\|_\infty^q \left(\mathbb{E}|Y_0|^{2q}\right)^{1/q} n^{-M_0/2}\left(\mathbb{E}\exp(|Y_0|)\right)^{1/2} \\ &\to 0 \ (n\to\infty),\end{aligned}$$

for $M(n) = M_0\log(n)$, $M_0 > 0$ big enough. It remains to check that $S_n = (\tilde{g}-\mathbb{E}\tilde{g})(x) = \frac{1}{nh}\sum_{t=1}^n W_t$, $W_t = T_t - \mathbb{E}T_t$, satisfies

$$\overline{\lim}_{n\to\infty}(nh)^{q/2}\,|\mathbb{E}S_n|^q < \infty.$$

We set for any integer $l \in [1,n]$,

$$A_l(\tilde{g}) := \sum_{1\le t_1\le\ldots\le t_l\le n} |\mathbb{E}W_{t_1}\ldots W_{t_l}|.$$

As in [7], the method relies on the inductive relationship,

$$A_q(\tilde{g}) \le n|\mathbb{E}W_0^q| + C_q n\sum_{r=1}^{n-1}(r+1)^{q-2}C_{r,q}(\tilde{g}) + \sum_{m=2}^{q-2}A_m(\tilde{g})A_{q-m}(\tilde{g}).$$

Here $C_{r,q} = C_{r,q}(\tilde{g})$ is defined as

$$C_{r,q} = \sup\{|\mathrm{Cov}(W_{t_1}...W_{t_m}, W_{t_{m+1}}...W_{t_q})|;\ 1 \le t_1 \le \ldots \le t_q \le n,\ t_{m+1} - t_m = r,\ 1 \le m \le q-1\},$$

which is a bound for the expressions

$$|\mathrm{Cov}(W_{t_1}...W_{t_m}, W_{t_{m+1}}...W_{t_q})|,$$

taken on the for all increasing sequences $(t_i)_{i=1}^q$ with the same length $r = t_{m+1} - t_m$ of the maximum spacing. Clearly,

$$n|\mathbb{E}W_0^q| \le \mathrm{C}\ nh\log^q(n). \tag{5.15}$$

The boundedness assumptions on the densities in (4.4) yield a first bound for $C_{r,q}$,

$$C_{r,q} \le \mathrm{C}\,(\log(n))^{q-2}\,h^2. \tag{5.16}$$

The $(\theta, \psi_1, \mathcal{L})$-weak dependence of the process yields a second bound for $C_{r,q}$,

$$C_{r,q} \le \mathrm{C}\frac{(\log(n))^q\,\theta_r}{h}. \tag{5.17}$$

We use again Tran's truncation technique. Assume that $q \ge 4$ is an even integer. If $a > \max(q-1, \frac{6\delta(q-1)}{q-2+\delta(4-q)})$, then there exists $\varsigma \in]0,1[$ such that $((2+q)\delta + 2 - q)/2(a-q+1) < \varsigma < ((4-q)\delta + q - 2)/2(q-2)$, so that

$$\begin{aligned}(nh)^{-q/2}\sum_{r=1}^{[n^\varsigma]} n\,(r+1)^{q-2}\,C_{r,q} &\le \mathrm{C}(nh)^{-q/2}nh^2\,(\log(n))^{q-2}\,[n^\varsigma]^{q-1} \\ &\le n^{(2\varsigma(q-1)-((4-q)\delta+q-2))/2}\,(\log(n))^{q-2} \to 0\ (n\to\infty),\end{aligned} \tag{5.18}$$

and

$$\begin{aligned}(nh)^{-q/2}\sum_{r=1+[n^\varsigma]}^{n-1} n\,(r+1)^{q-2}\,C_{r,q} &\le \mathrm{C}(nh)^{-q/2}nh^{-1}\,(\log(n))^{q}\,[n^\varsigma]^{q-1-a} \\ &\le n^{(-2\varsigma(a-q+1)+(q+2)\delta+2-q)/2}\,(\log(n))^{q} \to 0\ (n\to\infty).\end{aligned} \tag{5.19}$$

Formulae (5.15), (5.18) and (5.19) show that the numbers $B_l = (nh)^{-l/2}A_l(\tilde{g})$ satisfy the relation $B_q \le \sum_{m=2}^{q-2} B_m B_{q-m} + c_0$. Therefore, (see [4] for more details on Catalan numbers B_l) $B_q \le c_0\frac{(2q-2)!}{q!(q-1)!}$, that is

$$\overline{\lim}_{n\to\infty}(nh)^{q/2}\,|\mathbb{E}S_n|^q \le \overline{\lim}_{n\to\infty} q!\,(nh)^{-q/2}A_q(\tilde{g}) \le c_0,$$

which completes the proof for $q \ge 4$ an even integer and in case of a $(\theta, \psi_1, \mathcal{L})$-weak dependent process. The case when q is an odd integer ≥ 3 is analogous.

If the process is $(\theta, \psi_2, \mathcal{L})$-weak dependent, the second bound for $C_{r,q}$ becomes,

$$C_{r,q} \le \mathrm{C}\frac{(\log(n))^q\,\theta_r}{h^2},$$

in place of (5.17). □

Remark 5.2. The conclusion of Theorem 4.2 also holds true for a stationary, $(\theta, \psi, \mathcal{I})$-weak dependent process under the condition on the dependence rate $\theta_r = O(r^{1-q})$; see section 2.1 for the class $\mathcal{I}$. The results can also be obtained under a polynomial moment assumption as in (4.2). For instance, if $S > S(q,\delta) = \max\left(\frac{2q(q-2+\delta)}{(q-2)+\delta(4-q)}, \frac{2q\delta}{(q-2)(1-\delta)}\right)$, tedious computations show that the assertion of Theorem 4.2 holds true for a $(\theta, \psi_1, \mathcal{L})$-weak dependent process with $\theta_r = O(r^{-a})$, for some

$$a > a_1(q,\delta,S) = \max\left(q-1, \frac{3\,(q-1)\,(q+\delta(4S-3q))}{2\delta\,(4S-qS-2q)+2\,(q-2)\,(S-2q)}\right).$$

For a $(\theta, \psi_2, \mathcal{L})$-weak dependent process the assertion of Theorem 4.2 holds if

$$a > a_2(q,\delta,S) = \max\left(q-1, \frac{4\,(q-1)\,(q+\delta(4S-3q))}{2\delta\,(4S-qS-2q)+2\,(q-2)\,(S-2q)}\right).$$

5.3 Asymptotic normality of the regression estimator

Proof of Proposition 4.2. We follow [3] and [20]. From the expansion

$$u^{-1} = \sum_{i=0}^{p}(-1)^i\frac{(u-u_0)^i}{u_0^{i+1}} + (-1)^{p+1}\frac{(u-u_0)^{p+1}}{uu_0^{p+1}}, \tag{5.20}$$

we deduce that

$$\begin{aligned}\mathbb{E}\left(\hat{r}(x)\right) &= \mathbb{E}a_n - \mathbb{E}\left(a_n-\mathbb{E}a_n\right)\left(b_n-\mathbb{E}b_n\right) + \left(\mathbb{E}a_n\right)\mathbb{E}\left(b_n-\mathbb{E}b_n\right)^2 + \mathbb{E}\left(a_n-\mathbb{E}a_n\right)\left(b_n-\mathbb{E}b_n\right)^2 \\ &- \mathbb{E}\hat{r}(x)\left(b_n-\mathbb{E}b_n\right)^3, \end{aligned}\tag{5.21}$$

where, in equation (5.20), $p=2$, $u=b_n$, $u_0=\mathbb{E}b_n=1$, and $\hat{r}(x)=a_n/b_n$ (if $b_n\neq 0$) with

$$a_n = \sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h_n}\right) / \left(n\mathbb{E}K\left(\frac{x-X_0}{h_n}\right)\right), \text{ and } b_n = \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) / \left(n\mathbb{E}K\left(\frac{x-X_0}{h_n}\right)\right).$$

Clearly, since f and g are ρ-regular,

$$\mathbb{E}a_n = r(x) + O(h^\rho).$$

Assume that the sequence is $(\theta, \psi_1, \mathcal{L})$ weak-dependent. By the Cauchy-Schwarz inequality, we bound the second right hand side term in equation (5.21) with the variance bounds given in Proposition 4.1. It is of the order $O\left((nh)^{-1}\right)$ if $a>3$. The third term has the same order.

Rosenthal type results, stated in Theorem 4.2, then yield

$$\begin{aligned}\left|\mathbb{E}\left(a_n-\mathbb{E}a_n\right)\left(b_n-\mathbb{E}b_n\right)^2\right| &\leq \left(\frac{h}{\mathbb{E}K\left(\frac{x-X_0}{h_n}\right)}\right)^3 \left(\mathbb{E}\left(\hat{g}(x)-\mathbb{E}\hat{g}(x)\right)^4\right)^{1/2}\left(\mathbb{E}\left(\hat{f}(x)-\mathbb{E}\hat{f}(x)\right)^2\right)^{1/2} \\ &= O\left((nh)^{-3/2}\right)\end{aligned}$$

as soon as $a > a_1(4,\delta)$.

For the last term, we use a truncation device

$$\begin{aligned}\left(\mathbb{E}\left(\hat{r}(x)\right)^4\right)^{1/4} &\leq M(n) + \left(\mathbb{E}\left(\frac{\sum_{i=1}^n Y_i^4 \mathbb{I}_{\{|Y_i|>M(n)\}} K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)}\right)\right)^{1/4} \\ &= O\left(\log(n) + n^{(2-M_0)/8}\right)\end{aligned}$$

if $M(n) = M_0 \log(n)$. Thus, by Hölder's inequality,

$$\mathbb{E}\hat{r}(x)\left(b_n - \mathbb{E}b_n\right)^3 = O\left(\log(n)(nh)^{-3/2}\right)$$

if $a > a_1(4,\delta)$. This completes the proof in case of $(\theta,\psi_1,\mathcal{L})$ weak-dependence. The $(\theta,\psi_2,\mathcal{L})$ weak-dependent case is quite similar, the details are omitted. □

Proof of Proposition 4.3. We use Collomb's expansion (5.20). Here $p=1$, $u=\hat{f}(x)$, $u_0=\mathbb{E}\hat{f}(x)$. Thus,

$$\begin{aligned}&\hat{r}(x) - \mathbb{E}\hat{r}(x) \\ = &\frac{\hat{g}(x)\mathbb{E}\hat{f}(x) - \hat{f}(x)\mathbb{E}\hat{g}(x)}{\left(\mathbb{E}\hat{f}(x)\right)^2} + \frac{\left(\mathbb{E}\hat{g}(x) - \hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right) + \mathbb{E}\left(\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)\right)}{\left(\mathbb{E}\hat{f}(x)\right)^2} \\ + &\frac{\hat{r}(x)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2 - \mathbb{E}\left(\hat{r}(x)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2\right)}{\left(\mathbb{E}\hat{f}(x)\right)^2}.\end{aligned}$$

Now,

$$\mathrm{Var}\left(\hat{r}(x)\right) = \frac{\mathbb{E}\left(\hat{g}(x)\mathbb{E}\hat{f}(x) - \hat{f}(x)\mathbb{E}\hat{g}(x)\right)^2}{\left(\mathbb{E}\hat{f}(x)\right)^4} + o\left(\frac{1}{nh}\right) \tag{5.22}$$

under the conditions
$nh\mathbb{E}\left|\hat{g}(x) - \mathbb{E}\hat{g}(x)\right|^3 \leq \left((nh)^{4/3}\mathbb{E}\left|\hat{g}(x) - \mathbb{E}\hat{g}(x)\right|^4\right)^{3/4} \to 0$, $nh\mathbb{E}\left|\hat{r}(x)\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)\right|^3 \leq nh\left(\mathbb{E}\hat{r}^4(x)\right)^{1/4}$
$\left(\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^4\right)^{3/4} \to 0$, $nh\left(\mathrm{Var}(\hat{g}(x))\right)^2 \to 0$, and
$nh\left(\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^6\right)^{1/2}\left(\mathbb{E}\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)^6\right)^{1/6}\left(\mathbb{E}\left|\hat{r}(x)\right|^3\right)^{1/3} \to 0.$
The first bound is obtained if the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_j,\mathcal{L})$-weak dependent, for $j=1,2$, with $\theta_r = O(r^{-a})$, where

$$a > \max\left(3, \frac{9(j+2)\delta}{5-2\delta}, \frac{3(j+2)\delta}{1+2\delta}\right).$$

The second bound too. The third bound is satisfied if

$$a > \max\left(1, \frac{2(j+2)\delta}{1+\delta}\right).$$

The last bound holds true if

$$a > \max\left(5, \frac{10(j+2)\delta}{7-5\delta}\right).$$

The first right hand variance term is handled along the same lines than Proposition 4.1: $a > 2 + j$. □

Proof of Theorem 4.3. It is a straightforward application of the following identity,

$$\begin{aligned}\hat{r}(x) - r(x) &= \frac{\left(\hat{g} - r\hat{f}\right)(x) - \mathbb{E}\left(\left(\hat{g} - r\hat{f}\right)(x)\right)}{\mathbb{E}\hat{f}(x)} - \left(r(x) - \frac{\mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}(x)}\right) + \frac{\left(r(x) - \frac{\mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}(x)}\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)}{\mathbb{E}\hat{f}(x)} \\ &- \frac{(\hat{g}(x) - \mathbb{E}\hat{g}(x))\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)}{\left(\mathbb{E}\hat{f}(x)\right)^2} + \frac{\hat{r}(x)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2}{\left(\mathbb{E}\hat{f}(x)\right)^2}.\end{aligned}$$

Assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weak dependent, for $j = 1$ or $j = 2$. Then

$$\begin{aligned}\mathbb{E}\left|\sqrt{nh}\frac{\left(r(x) - \frac{\mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}(x)}\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)}{\mathbb{E}\hat{f}(x)}\right| &= O\left(nh^{(1+2\rho)}\right)^{1/2} \\ \mathbb{E}\left|\sqrt{nh}\frac{(\hat{g}(x) - \mathbb{E}\hat{g}(x))\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)}{\left(\mathbb{E}\hat{f}(x)\right)^2}\right| &\leq C\sqrt{nh}\left(\mathbb{E}\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)^2\right)^{1/2}\left(\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2\right)^{1/2} \\ &= O\left(nh\right)^{-1/2} \text{ if } a > 2 + j \\ \mathbb{E}\left|\sqrt{nh}\frac{\hat{r}(x)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2}{\left(\mathbb{E}\hat{f}(x)\right)^2}\right|^{1/2} &\leq (nh)^{1/4}\left(\mathbb{E}\left|\hat{r}(x)\right|\right)^{1/2}\left(\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2\right)^{1/2} \\ &= O\Big((nh)^{-1/4}\log^{1/2}(n)\Big) \text{ if } a > 2 + j\end{aligned}$$

Then the result follows from Theorem 4.1. □

Remark 5.3. The CLT for the centered regression function can be proved by the same device. In fact, consider the numbers $\alpha_1(\delta)$ and $\alpha_2(\delta)$ introduced in Remark 5.1. Define

$$\beta_1(\delta) = \max\left(\alpha_1(\delta), \max\Big(3, \frac{27\delta}{7 - 4\delta}\Big)\right) = \begin{cases} 3 & \text{if } 0 < \delta < \frac{1}{3}, \\ 9\delta & \text{if } \frac{1}{3} \leq \delta < \frac{1+\sqrt{5}}{6}, \\ 3 + \frac{1}{\delta} & \text{if } \frac{1+\sqrt{5}}{6} \leq \delta < \frac{17+\sqrt{1381}}{78}, \\ \frac{27\delta}{7-4\delta} & \text{otherwise,} \end{cases}$$

and

$$\beta_2(\delta) = \max\left(\alpha_2(\delta), \max\Big(3, \frac{36\delta}{7 - 4\delta}\Big)\right) = \begin{cases} 4 & \text{if } 0 < \delta < \frac{1}{3}, \\ 12\delta & \text{if } \frac{1}{3} \leq \delta < \frac{1}{2}, \\ 4 + \frac{1}{\delta} & \text{if } \frac{1}{2} \leq \delta < \frac{24+\sqrt{2032}}{104}, \\ \frac{36\delta}{7-4\delta} & \text{otherwise.} \end{cases}$$

Proposition 5.2 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the conditions (4.1) with $p = 2$, (4.3) and (4.4). Let $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. In addition, assume that the sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weak dependent, for $j = 1$ or $j = 2$, and $\theta_r = O(r^{-a})$, where $a > \beta_j(\delta)$. Then*

$$\sqrt{nh}\ \left(\hat{r}(x) - \mathbb{E}\hat{r}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{G_2(x)}{f^2(x)}\int K^2(u)du\right).$$

The proof of Proposition 5.2 is based upon the Collomb's expansion,

$$\begin{aligned}
&\hat{r}(x) - \mathbb{E}\hat{r}(x) \\
= \; & \frac{\hat{g}(x)\mathbb{E}\hat{f}(x) - \hat{f}(x)\mathbb{E}\hat{g}(x)}{\left(\mathbb{E}\hat{f}(x)\right)^2} + \frac{\left(\mathbb{E}\hat{g}(x) - \hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right) + \mathbb{E}\left(\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)\right)}{\left(\mathbb{E}\hat{f}(x)\right)^2} \\
+ \; & \frac{\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2 - \mathbb{E}\left(\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2\right) + \left(\mathbb{E}\hat{g}(x)\right)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2}{\left(\mathbb{E}\hat{f}(x)\right)^3} \\
- \; & \frac{\left(\mathbb{E}\hat{g}(x)\right)\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2}{\left(\mathbb{E}\hat{f}(x)\right)^3} + \frac{\hat{r}(x)\left(\mathbb{E}\hat{f}(x) - \hat{f}(x)\right)^3 + \mathbb{E}\left(\hat{r}(x)\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^3\right)}{\left(\mathbb{E}\hat{f}(x)\right)^3}.
\end{aligned}$$

The first right hand term is handled as in the proof of Proposition 5.1. The other terms are shown to be $o_p\left(\frac{1}{\sqrt{nh}}\right)$ by using relevant moments inequalities for $\hat{f}$ and $\hat{g}$. More precisely, we prove that $(nh)^{1/2}\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2 \to 0$ and $(nh)^{1/2}\mathbb{E}\left(\hat{g}(x) - \mathbb{E}\hat{g}(x)\right)^2 \to 0$, as $n \to \infty$, if the process $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weak dependent with

$$a > \max\left(1, \frac{2(2+j)\delta}{1+\delta}\right).$$

And $(\log(n))^{4/3}(nh)^{2/3}\mathbb{E}\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^4 \to 0$ if

$$a > \max\left(3, \frac{9(2+j)\delta}{7-4\delta}, \frac{3(2+j)\delta}{2+\delta}\right).$$

5.4 Almost sure convergence

Proof of Theorem 4.4. We keep the notations as in (5.1) and (5.2) and denote again by C a universal constant (whose value might change). Since $\mathbb{E}(\exp(|Y_0|)) < \infty$,

$$\mathbb{P}\left(\sup_{|x|\le M} |\hat{g} - \tilde{g}|(x) > 0\right) \le n\mathbb{P}\left(|Y_0| M_0 \log(n)\right) \le \mathrm{C} n^{1-M_0},$$

and, by the Cauchy-Schwarz inequality,

$$\sup_{|x|\le M} \mathbb{E}\left(|\hat{g} - \tilde{g}|(x)\right) \le \frac{1}{h}\,\mathbb{E}\left[|Y_0| \mathbb{1}_{\{|Y_0| M_0 \log(n)\}} |K\left(\frac{x - X_0}{h}\right)|\right] \le \frac{1}{h} h^{1/3} n^{-M_0}.$$

We can now reduce the computations to those of a density estimator, as given in [7]. Assume that the interval $[-M, M]$ is covered by L_ν intervals with diameter $1/\nu$ (here $\nu = \nu(n)$ depends on n); we denote by I_j the j-th interval and x_j its center. Assume that the relation $h\nu \to \infty$ holds (for $n \to \infty$). Then [14] proves that there exists another kernel-type density estimate $\tilde{g}'$ based on an even, continuous kernel, decreasing on $[0, \infty[$, constant on $[0, 2R_0]$, taking the value 0 at $t = 3R_0$, where the kernel K vanishes if $t > R_0$. Then,

$$\sup_{x\in I_j} |\tilde{g} - \mathbb{E}\tilde{g}|(x) \le |\tilde{g}(x_j) - \mathbb{E}\tilde{g}(x_j)| + \frac{\mathrm{C}}{h\nu}\left(|\tilde{g}' - \mathbb{E}\tilde{g}'|(x_j) + 2|\mathbb{E}\tilde{g}'|(x_j)\right).$$

Therefore, for any $\lambda > 0$,

$$\begin{aligned} & \mathbb{P}\Big(\sup_{x\in[-M,M]} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| \geq \frac{2\lambda}{\sqrt{nh}} + \frac{1}{h} h^{1/3} n^{-M_0} + \mathrm{C}\ \frac{\log(n)}{h\nu} \Big) \\ \leq\ & n^{1-M_0} + L_\nu \mathbb{P}\left(|\tilde{g} - \mathbb{E}\tilde{g}|(x_1) | \frac{\lambda}{\sqrt{nh}} \right) + L_\nu \mathbb{P}\left(|\tilde{g}' - \mathbb{E}\tilde{g}'|(x_1) \frac{\lambda}{\sqrt{nh}} \right). \end{aligned}$$

Following the same ideas as in [7], the proof of assertion **(i)** will be complete with the inequality in Lemma 5.1 below.

The proof of assertion **(ii)** in Theorem 4.4 is then straightforward, using Collomb's expansion in (5.20). □

Lemma 5.1 *Under the conditions of the Theorem 4.4, there exist positive constants F, G such that for any $\lambda > 0$ and any bandwidth $h \to 0$ with $nh \geq 1$,*

$$\sup_{x\in\mathbb{R}} \mathbb{P}\left(|\tilde{g}(x) - \mathbb{E}\tilde{g}(x)| > \frac{\lambda}{\sqrt{nh}} \right) \leq F \exp(-G\sqrt{\lambda}).$$

The same exponential inequality holds for $\tilde{g}'$, too.

Sketch of the proof of Lemma 5.1. We specify the constants in the covariance inequalities (5.15), (5.16) and (5.17) for a $(\theta, \psi_1, \mathcal{L})$-weak dependent process. Denote by $K_\infty = 2\|K\|_\infty$

$$C_1(q) = 2\|f\|_\infty R_0 K_\infty^q,\ C_2(q) = K_\infty^{q-2} \|f_{0,r}\|_\infty \left(\int |K(u)|\ du \right)^2,\ \text{and } C_3(q) = q^2 K_\infty^q Lip(K)/2.$$

Then for some constant $D > 0$, which neither depends on r nor on q,

$$(nh)^{-q/2} \sum_{r=1}^{n} n\,(r+1)^{q-2}\, C_{r,q} \leq (qD)^q\,.$$

The last estimations, together with the previous inductive relation on $A_q(g)$ yield

$$A_q(\tilde{g}) \leq \sum_{m=2}^{q-2} A_m(\tilde{g}) A_{q-m}(\tilde{g}) + (qD)^q\,.$$

which holds for all integers $q \geq 2$. Thus, the numbers $B_l = \frac{A_l(\tilde{g})}{\left(Dl\sqrt{nh}\right)^l}$ satisfy the relation $B_q \leq \sum_{m=2}^{q-2} B_m B_{q-m} + 1 \leq \frac{(2q-2)!}{q!(q-1)!}$. Therefore,

$$\begin{aligned} \mathbb{P}\left(|\tilde{g}(x) - \mathbb{E}\tilde{g}(x)| \geq \frac{\lambda}{\sqrt{nh}} \right) &\leq \lambda^{-q} q!\, A_q(\tilde{g})\, (nh)^{-q/2} \\ &\leq \left(\frac{q^2}{De\lambda} \right)^q \leq \exp\left(-D\sqrt{\lambda}\ \varphi\left(\frac{q}{\sqrt{D\lambda}} \right) \right) \end{aligned}$$

where $\varphi(t) = t\log(t)$. Now by optimizing the value of $\varphi(t)$ in the Markov inequality above, we obtain the exponential inequality (see [7] or [9] for more details). □ .

We also obtain a result under a polynomial moment assumptions. The two weak dependence conditions lead to quite similar results. We restrict ourselves to the $(\theta, \psi_1, \mathcal{L})$ case.

Proposition 5.3 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_1,\mathcal{L})$-weak dependent and satisfies the conditions (4.1) with $p=2$ and (4.4). Assume that $q\geq 2$ is an even integer, $f,g\in\mathcal{C}_\rho$ for some $\rho\in]0,\infty[$. In addition, assume both of the following.*

(i) *Assumption (4.2) holds, with $S>S_1(\rho,q)\ \in\ ]2q,3q[$, $S_1(\rho,q)=2q+\epsilon_1(\rho,q)$, where $\epsilon_1(\rho,q)\to 0$ as $\rho\to\infty$,*

(ii) *$\theta_r=O(r^{-a})$ with $a>q-1+\epsilon_2(\rho,q,S)$, where $\epsilon_2(\rho,q,S)\to 0$ as $\rho\to\infty$.*

If $\inf_{|x|\leq M} f(x)>0$, and $n^\delta h\to\infty$ for some $\delta\in]0,\frac{S(q-2)(1+2\rho)}{2+S(2+2\rho+q(1+2\rho))}[$, then, almost surely,

$$\sup_{|x|\leq M}|\hat{r}(x)-r(x)|=O\left(n^{-\frac{\rho S(q-2)}{2+S(2+2\rho+q+2q\rho)}}\right).$$

Remark 5.4. Zhao and Fang [29] prove almost sure convergence, uniformly on compact sets, of the kernel regression estimator for strongly mixing stationary process under the same condition as in Theorem 4.4. Let us consider a strongly mixing processes (see [1]) that satisfies the conditions (4.1) with $p=2$ and (4.4). Assume that $f,g\in\mathcal{C}_\rho$ for some $\rho\in]0,\infty[$ and $n^\delta h\to\infty$ for some $\delta\in]0,1[$. If the moment condition (4.2) holds with $S>4+\frac{2}{\rho}$, and if the mixing rate is $\alpha_r\leq r^{-a}$ for some $a>4$ such that

$$a>\max\left(\frac{2S}{S-2},\frac{S(S+4)(1+\rho)}{(S-2)(S\rho-4\rho+4)},\frac{(S+1)(3+4\rho)}{\rho S-2-4\rho}\right),$$

then, almost surely,

$$\sup_{|x|\leq M}|\hat{r}(x)-r(x)|=O\left((n^{-1}\log(n))^{\rho/(1+2\rho)}\right).$$

Sketch of the proof of Proposition 5.3. The strategy is the same as in the proof of Theorem 4.4. We deduce from the Rosenthal inequality in section 4.2, Remark 4.3, that for suitable constants

$$\sup_{x\in\mathbb{R}}\mathbb{P}\left(|\hat{g}(x)-\mathbb{E}\hat{g}(x)|\geq\frac{\lambda}{\sqrt{nh}}\right)\leq\frac{\mathrm{C}}{\lambda^q}.$$

Then, using the decomposition device of [14],

$$\mathbb{P}\left(\sup_{x\in[-M,M]}|\tilde{g}(x)-\mathbb{E}\tilde{g}(x)|\geq\frac{2\lambda}{\sqrt{nh}}+\frac{c}{\nu h^{(1+1/S)}}\right)\leq\frac{C\ \nu\ (1+(h\nu)^{-q})}{\lambda^q}.$$

The term λ is chosen for balancing the bias term h^ρ. Appropriate choices of ν and S will complete the proof by using Borel-Cantelli's Lemma. □

5.5 Local bootstrap

Proof of Theorem 4.5. We follow the proof of Theorem 2.1 (i) in [19]. Their conditions (A3) and (A8) hold by assumption, the Lipschitz property of the kernel $K(.)$ with compact support is sufficient for the part in their assumption (A4) which is used to achieve Theorem 2.1 (i) in [19]. Asymptotic normality in (A6) of [19] holds by our Proposition 5.2. Then, Lemma 5.1 for $\varphi(x)=x^s$ $(s=1,2,3,4)$

in [19] follows from our Theorem 4.4 **(ii)**, see also Remark 4.4. All what remains to do for assertion **(i)** is to check the T_1 term in the proof of Lemma 5.2. in [19]. In particular, it suffices to show

$$(nh)^{-1}\sum_{t=1}^{n} V(X_t)K^2\left(\frac{x-X_t}{h}\right) = f(x)V(x)\int K^2(w)dw + o_P(1)\ (n\to\infty), \tag{5.23}$$

$$(nh)^{-1}\sum_{t=1}^{n} M(X_t)K^4\left(\frac{x-X_t}{h}\right) = f(x)M(x)\int K^4(w)dw + o_P(1)\ (n\to\infty). \tag{5.24}$$

Let us consider first (5.23). Since we do not assume that V is in $\mathcal{L}$, i.e. a bounded Lipschitz function, there is not a direct way to achieve (5.23). By the compact support of K, the term $V(X_t)K((x-X_t)/h)$ is zero for X_t outside a neighborhood of x and if n is sufficiently large. Thus, consider a modification $\tilde{V}(.)$ which is bounded, Lipschitz, satisfies (4.5) and with the requirement that $\tilde{V}(x) = V(x)$ at the point x. Then, write the left hand side of (5.23) as

$$\begin{aligned}(nh)^{-1}\sum_{t=1}^{n} V(X_t)K^2\left(\frac{x-X_t}{h}\right) &= (nh)^{-1}\sum_{t=1}^{n} \tilde{V}(X_t)K^2\left(\frac{x-X_t}{h}\right) + \Delta_n,\\ \Delta_n &= (nh)^{-1}\sum_{t=1}^{n}(\tilde{V}(X_t) - V(X_t)K^2\left(\frac{x-X_t}{h}\right).\end{aligned} \tag{5.25}$$

By (4.5) applied to $\tilde{V}(.)$, the fact that $\tilde{V}(x) = V(x)$ and the continuity of $f(.)$,

$$\mathbb{E}\left((nh)^{-1}\sum_{t=1}^{n} \tilde{V}(X_t)K^2\left(\frac{x-X_t}{h}\right)\right) = f(x)V(x)\int K^2(w)dw + o(1). \tag{5.26}$$

Moreover, since $\tilde{V}(.)$ and $K(.)$ are bounded Lipschitz functions it follows from the weak-dependence assumption that

$$\mathrm{Var}\left((nh)^{-1}\sum_{t=1}^{n} \tilde{V}(X_t)K^2\left(\frac{x-X_t}{h}\right)\right) = o(1)\ (n\to\infty). \tag{5.27}$$

Finally,

$$\begin{aligned}\mathbb{E}|\Delta_n| &\le \mathbb{E}\left((nh)^{-1}\sum_{t=1}^{n}|V(X_t) - \tilde{V}(X_t)|K^2\left(\frac{x-X_t}{h}\right)\right)\\ &= \int |V(v) - \tilde{V}(v)|K^2\left(\frac{x-v}{h}\right)f(v)dv = O(h^2),\end{aligned} \tag{5.28}$$

where the last bound follows by assumption (4.5) for $V(.)$ and $\tilde{V}(.)$. By (5.25)-(5.28) we have shown (5.23). The proof of (5.24) is analogous. This then completes the proof of the first assertion.

The second assertion is immediate. □

References

[1] Ango Nze, P., Doukhan, P. (1998). Functional estimation for time series: Uniform Convergence Properties. J. Statist. Plann. Infer. 68, 5-29.

[2] Bickel, P.J., Bühlmann, P. (1999). A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. Bernoulli 5, 413-446.

[3] Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau. Zeit. Wahrsch. verw. Gebiete 66, 441-460.

[4] Comtet, L. (1970). *Analyse combinatoire*, tome 1, P.U.F., Masson, Paris.

[5] Coulon, C., Doukhan, P. (1999). A triangular central limit theorem under a weak dependence condition. *To appear in Statist. & Probab. Lett.*

[6] Doukhan, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics 85.* Springer Verlag.

[7] Doukhan, P., Louhichi, S. (1997). Weak dependence and moment inequalities. Université de Paris-Sud, Prépublication 97.08.

[8] Doukhan, P., Louhichi, S. (1998). Functional estimation of a density under a new weak dependence condition. C. R. Acad. Sci. Paris, série 1, 327, 989-992.

[9] Doukhan, P., Portal, F. (1987). Principe d'invariance faible pour la fonction de répartition empirique dans un cadre multidimensionnel et mélangeant. Probab. Math. Statist. 8, 117-132.

[10] Duflo, M. (1996). *Algorithmes stochastiques.* Math. Appl. 23. Springer Verlag.

[11] Esary, J., Proschan, F., Walkup, D. (1967). Association of random variables with applications. Ann. Math. Statist. 38, 1466-1476.

[12] Franke, J., Kreiss, J.-P., Mammen, E., Neumann, M.H. (1998). Properties of the nonparametric autoregressive bootstrap. Preprint.

[13] Hart, J.D. (1995). Some automated methods of smoothing time-dependent data. J. Nonpar. Statist. 6, 115-142.

[14] Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. Stoch. Process. Appl. 65, 69-80.

[15] Mokkadem, A. (1990). Propriétés de mélange des modèles autorégressifs polynômiaux. Ann. Inst. H. Poincaré B, 26, 219-260.

[16] Neumann, M.H., Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregression. Ann. Statist. 26, 1570-1613.

[17] Neumann, M.H. (1998). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. Ann. Statist. 26, 2014-2048.

[18] Newman, C.M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. Inequalities in Statist. Probab. 5, 127-140.

[19] Paparoditis, E., Politis, D.N. (1998). The local bootstrap for kernel estimators under general dependence conditions. *To appear in Ann. Inst. Statist. Math.*

[20] Pawlak, M. (1991). On the almost everywhere properties of the kernel regression estimate. Ann. Inst. Statist. Math. 43, 311-326.

[21] Prakasa, R. (1983). *Nonparametric functional estimation.* Academic Press.

[22] Rio, E. (1996). Sur le théorème de Berry-Esseen pour les suites faiblement dépendantes. Prob. Theory Relat. Fields 104, 255-282.

[23] Rio, E. (1999). Théorèmes limites pour les suites de variables aléatoires faiblement dépendantes. *To appear.*

[24] Robinson, P.M. (1983). Nonparametric estimators for time series. J. Time Ser. Anal. 4, 185-207.

[25] Rosenblatt, M. (1980). Linear Processes and bispectra. J. Appl. Prob. 17, 265-270.

[26] Rosenblatt, M. (1991). *Stochastic curve estimation.* NSF-CBMS Regional Conference Series in Probability and Statistics, 3.

[27] Tong, H. (1981). A note on a Markov bilinear stochastic process in discrete time. J. Time Ser. Anal. 2, 279-284.

[28] Tran, L. (1990). Recursive kernel density estimators under a weak dependence condition. Ann. Inst. Statist. Math. 42, 305-329.

[29] Zhao, L.C., Fang, Z. (1985). Strong convergence of kernel estimates of nonparametric regression functions. Chin. Ann. Math. B, 6, 147-155.

Université Lille 3
E.A. GREMARS, UFR AES, BP 149
59653 Villeneuve d'Ascq Cedex
France
E-mail: angonze@univ-lille3.fr

ETH Zürich, Seminar für Statistik
CH-8092 Zürich
Switzerland
E-mail: buhlmann@stat.math.ethz.ch

Université Cergy Pontoise
J.E. 2070 Mathematics, Dept. of Economy
33 Boulevard du Port
95011 Cergy-Pontoise Cedex
France
E-mail: doukhan@math.u-cergy.fr