Ann Inst Stat Math (2019) 71:1233–1263
https://doi.org/10.1007/s10463-018-0674-9

On the strong universal consistency of local averaging regression estimates

Matthias Hansmann[1] · **Michael Kohler**[1] · **Harro Walk**[2]

Received: 14 February 2018 / Revised: 21 May 2018 / Published online: 30 June 2018

Abstract A general result concerning the strong universal consistency of local averaging regression estimates is presented, which is used to extend previously known results on the strong universal consistency of kernel and partitioning regression estimates. The proof is based on ideas from Etemadi's proof of the strong law of large numbers, which shows that these ideas are also useful in the context of strong laws of large numbers for conditional expectations in L_2.

Keywords Regression estimation · Strong universal consistency · Local averaging estimates · L_2 error

1 Introduction

Let $(X, Y), (X_1, Y_1), (X_2, Y_2), \ldots$ be independent identically distributed $\mathbb{R}^d \times \mathbb{R}$-valued random vectors with $\mathbf{E}Y^2 < \infty$. In regression analysis we want to estimate Y after having observed X, i.e., we want to determine a function f with $f(X)$ "close" to Y. If "closeness" is measured by the mean squared error, then one wants to find a function f^* such that

✉ Michael Kohler
kohler@mathematik.tu-darmstadt.de

Matthias Hansmann
hansmann@mathematik.tu-darmstadt.de

Harro Walk
walk@mathematik.uni-stuttgart.de

1 Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany

2 Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany

$$\mathbf{E}\left\{\left|f^*(X) - Y\right|^2\right\} = \min_f \mathbf{E}\left\{|f(X) - Y|^2\right\}. \tag{1}$$

Let $m(x) := \mathbf{E}\{Y|X = x\}$ be the regression function and denote the distribution of X by μ. The well-known relation which holds for each measurable function f

$$\mathbf{E}\left\{|f(X) - Y|^2\right\} = \mathbf{E}\{|m(X) - Y|^2\} + \int | f(x) - m(x) |^2 \mu(\mathrm{d}x) \tag{2}$$

(cf., e.g., Section 1.1 in Györfi et al. 2002) implies that m is the solution of the minimization problem (1), $\mathbf{E}\{|m(X) - Y|^2\}$ is the minimum of (2), and for an arbitrary f, the L_2 error $\int | f(x) - m(x) |^2 \mu(\mathrm{d}x)$ is the difference between $\mathbf{E}\{|f(X) - Y|^2\}$ and $\mathbf{E}\{|m(X) - Y|^2\}$.

In the regression estimation problem the distribution of (X, Y) (and consequently m) is unknown. Given a sequence $\mathcal{D}_n = \{(X_1, Y_1),\dots,(X_n, Y_n)\}$ of independent observations of (X, Y), our goal is to construct an estimate $m_n(x)=m_n(x, \mathcal{D}_n)$ of $m(x)$ such that the L_2 error $\int|m_n(x)-m(x)|^2\mu(\mathrm{d}x)$ is small, i.e., by choice $f = m_n$ the minimum in (1) is nearly attained.

A sequence of estimators $(m_n)_{n\in\mathbb{N}}$ is called **weakly universally consistent** if

$$\mathbf{E}\int|m_n(x) - m(x)|^2 \mu(\mathrm{d}x) \to 0 \quad (n \to \infty)$$

for all distributions of (X, Y) with $\mathbf{E}Y^2 < \infty$. It is called **strongly universally consistent** if

$$\int|m_n(x) - m(x)|^2 \mu(\mathrm{d}x) \to 0 \quad (n \to \infty) \quad a.s.$$

for all distributions of (X, Y) with $\mathbf{E}Y^2 < \infty$.

Stone (1977) first pointed out that there exist weakly universally consistent estimators. He considered k_n-*nearest neighbor estimates*

$$m_n(x) = \sum_{i=1}^{n} W_{n,i}(x) \cdot Y_i \tag{3}$$

where

$$W_{n,i}(x) = W_{n,i}(x, X_1, \dots, X_n) \tag{4}$$

is $1/k_n$ if X_i is among the k_n-nearest neighbors of x in $\{X_1, \dots, X_n\}$ and zero otherwise, and where $k_n \to \infty$ and $k_n/n \to 0$ $(n \to \infty)$. The strong universal consistency of nearest neighbor estimates has been shown in Devroye et al. (1994).

Estimates of the form (3) with weight functions (4) are called local averaging estimates. As a basis for his result on nearest neighbor estimates, Stone (1977) established a general theorem giving sufficient (and in some sense also necessary) conditions on weak universal consistency of local averaging estimates.

The most popular examples of local averaging estimates are the *Nadaraya–Watson kernel estimates* (Nadaraya 1964; Watson 1964), where

$$W_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)} \tag{5}$$

($0/0 = 0$ by definition) for some function $K : \mathbb{R}^d \to \mathbb{R}_+$ (called kernel) and some $h_n > 0$ (called bandwidth) usually satisfying

$$h_n \to 0 \quad (n \to \infty) \quad \text{and} \quad n \cdot h_n^d \to \infty \quad (n \to \infty), \tag{6}$$

e.g.,

$$h_n = const \cdot n^{-\gamma} \quad \text{for some } \gamma \in \mathbb{R} \text{ with } 0 < \gamma \cdot d < 1. \tag{7}$$

Frequently used kernels are the naive kernel (window kernel)

$$K(x) = I_{\{\|x\| \le 1\}} \quad (x \in \mathbb{R}^d)$$

(where I_A denotes the indicator function of a set A, and $\|x\|$ is the Euclidean norm of $x \in \mathbb{R}^d$), the Epanechnikov kernel

$$K(x) = \prod_{l=1}^d K_0(x^{(l)}) \quad (x = (x^{(1)}, \ldots, x^{(d)}) \in \mathbb{R}^d)$$

with

$$K_0(x^{(l)}) = \left(1 - |x^{(l)}|^2\right) \cdot I_{\{|x^{(l)}| \le 1\}}$$

(Epanechnikov 1969, with a discussion of an optimality property after some standardization) and the Gaussian kernel

$$K(x) = \exp\left(-\|x\|^2\right) \quad (x \in \mathbb{R}^d).$$

Another example for a local averaging estimate is *partitioning regression estimation* (regressogram introduced by Tukey 1947).

Weak universal consistency of kernel estimates has been shown for so-called boxed kernels, i.e., kernels which satisfy

$$c_1 \cdot I_{\{\|x\| \le r_1\}} \le K(x) \le c_2 \cdot I_{\{\|x\| \le r_2\}} \quad (x \in \mathbb{R}^d)$$

for some $c_1, c_2, r_1, r_2 > 0$ (e.g., the naive kernel or the Epanechnikov kernel) and bandwidths satisfying (6) independently by Devroye and Wagner (1980) and Spiegelman and Sacks (1980). Here in the second paper a slight modification of (5) is used, where the original denominator is replaced by the maximum of 1 and the original denominator. Strong universal consistency of kernel estimates for the naive kernel and suitably defined piecewise constant sequences of bandwidths has been shown by Walk (2002). Various results concerning strong universal consistency of variants of kernel estimates can be found in Györfi and Walk (1996, 1997) and Györfi et al. (1998). Walk (2005) treated smooths kernels, e.g., the Gaussian kernel, using the Spiegelman–Sacks modification of (5). Corresponding results for partitioning estimation are obtained by Györfi (1991), by Györfi et al. (1998) and in Section 23.1 in Györfi et al. (2002). Results concerning strong universal consistency of various least squares estimates are presented in Lugosi and Zeger (1995) and Kohler (1997, 1999). Kohler and Krzyżak (2001) and Kohler (2003) showed the universal consistency of suitably defined penalized least squares estimates and local polynomial kernel regression estimates, respectively. Further references can be found in Györfi et al. (2002). Related results in connection with strong (universal) pointwise consistency can be found in Devroye (1981), Greblicki et al. (1984), Devroye and Krzyżak (1989), Irle (1997), Kozek et al. (1998), Walk (2001, 2008) and Biau and Devroye (2015).

Surprisingly, despite the many existing results in the literature on strong universal consistency, there is still a gap in the literature concerning the strong universal consistency of the classical kernel regression estimate with weights (5). The only known result in this context (Walk 2002) requires that the kernel is the naive kernel and that the sequence of bandwidths is piecewise constant. The purpose of this paper is to fill this gap. In particular, we show that the classical kernel regression estimate with a boxed kernel (e.g., the Epanechnikov kernel) and with a sequence of bandwidths satisfying

$$h_n \downarrow 0 \quad (n \to \infty) \quad \text{and} \quad n \cdot h_n^d \uparrow \infty \quad (n \to \infty)$$

(e.g., the bandwidths defined in (7)) is strongly universally consistent. Thus, essentially the general result of Devroye and Wagner (1980) is sharpened from weak to strong universal consistency.

To achieve this result, at first a general theorem of Stone type on strong universal consistency of local averaging estimates with weights satisfying

$$\sum_{i=1}^{n} |W_{n,i}(x)| \leq D \quad (x \in \mathbb{R}^d)$$

for some $D \geq 1$ is presented (Theorem 1). Because strong universal consistency of nearest neighbor estimates is known (Devroye et al. 1994), only applications of this result to kernel estimates and partitioning estimates are presented. The application to kernel estimates (Theorem 2) yields the above-described result for boxed kernels. And the application to partitioning estimates (Theorem 3) sharpens the general weak universal consistency result on partitioning estimation of Theorem 4.2 in Györfi et al. (2002) to strong universal consistency in the case of nested partitioning (where the sets in the partitions are subsequently refined). In both applications the crucial step is the verification of the first condition in the strong consistency theorem, which allows the extension from bounded Y to square integrable Y. An essential tool is here the idea of “thinning” via majorization developed by Etemadi (1981) for proving the strong law of large numbers, together with refined covering arguments. The structure of kernel and partitioning weights allows a direct treatment of variances via binomial and multinomial distributions, avoiding the use of the Efron–Stein inequality (cf., e.g., Györfi et al. 2002, Theorem A.3), which seems to be too rough.

Throughout this paper we use the following notation: $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{R}_+$ are the sets of positive integers, real numbers and nonnegative real numbers, respectively. For $z \in \mathbb{R}$ we denote the smallest integer greater than or equal to z by $\lceil z \rceil$. $\|x\|$ is the Euclidean norm of $x \in \mathbb{R}^d$. For $z > 0$, $\log(z)$ is the natural logarithm and $\log_2(z)$ the logarithm with basis 2. For $A \subseteq \mathbb{R}^d$ we denote the diameter of A by

$$\operatorname{diam}(A) = \sup_{x,z \in A} \|x - z\|,$$

and we let I_A be the indicator function corresponding to A (which is one on A and zero otherwise). And for $x \in \mathbb{R}^d$ and $h \in \mathbb{R}$ we set

$$x + h \cdot A = \{x + h \cdot z \,:\, z \in A\}.$$

S_r denotes a closed Euclidean ball in $\mathbb{R}^d$ centered at 0 with radius $r > 0$. If $\mathcal{P}$ is a partition of $\mathbb{R}^d$ and $x \in \mathbb{R}^d$, then $A_{\mathcal{P}}(x)$ is the unique set $A \in \mathcal{P}$ with $x \in A$.

The outline of this paper is as follows: The main results are presented in Sect. 2 and proven in Sect. 3.

2 Main results

The following theorem is a tool for proving strong consistency for local averaging regression function estimates and corresponds to Stone’s (1977) general theorem on weak consistency. It will be applied to establish strong universal consistency of boxed kernel and nested partitioning regression estimates (Theorems 2 and 3).

Theorem 1 *Let (X, Y), (X_1, Y_1), (X_2, Y_2), …be independent and identically distributed $\mathbb{R}^d \times \mathbb{R}$ valued random vectors. Assume that the weights*

$$W_{n,i}(x) = W_{n,i}(x, X_1, \ldots, X_n)$$

of the local averaging regression estimate

$$m_n(x) = \sum_{i=1}^{n} W_{n,i}(x) \cdot Y_i$$

satisfy the following five conditions:

(A1) There exists a constant $c > 0$ *such that we have for every distribution of* Y *with* $Y \geq 0$ *a.s. and* $\mathbf{E}Y < \infty$

$$\limsup_{n\to\infty} \int \sum_{i=1}^{n} |W_{n,i}(x)| \cdot Y_i \, \mu(dx) \leq c \cdot \mathbf{E}Y \quad a.s.$$

(A2) For all $\delta > 0$*:*

$$\int \sum_{i=1}^{n} |W_{n,i}(x)| \cdot I_{\{\|X_i - x\| > \delta\}} \mu(dx) \to 0 \quad a.s.$$

(A3)

$$\log(n) \sum_{i=1}^{n} \left(\int |W_{n,i}(x)| \mu(dx) \right)^2 \to 0 \quad a.s.$$

(A4)

$$\int \left| \sum_{i=1}^{n} W_{n,i}(x) - 1 \right| \mu(dx) \to 0 \quad a.s.$$

(A5) There exists $D \geq 1$ *such that for all n and for* μ*—almost all* $x \in \mathbb{R}^d$*:*

$$\sum_{i=1}^{n} |W_{n,i}(x)| \leq D \quad a.s.$$

Then $(m_n)_{n\in\mathbb{N}}$ *is strongly universally consistent.*

Theorem 2 *Let* $K : \mathbb{R}^d \to \mathbb{R}$ *be a boxed kernel, i.e., assume that* K *satisfies*

$$c_1 \cdot I_{S_{r_1}}(x) \leq K(x) \leq c_2 \cdot I_{S_{r_2}}(x) \quad (x \in \mathbb{R}^d)$$

for some $c_1, c_2, r_1, r_2 > 0$*. Let* $h_n > 0$ *be such that*

$$h_n \to 0 \quad (n \to \infty) \tag{8}$$

$$n \cdot h_n^d \to \infty \quad (n \to \infty) \tag{9}$$

and

$$\limsup_{n\to\infty} \frac{\max\{h_n, h_{n+1}, \ldots, h_{2n}\}}{\min\{h_n, h_{n+1}, \ldots, h_{2n}\}} < \infty. \tag{10}$$

Define the kernel regression estimate by

$$m_n(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \cdot Y_i}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)}.$$

Then $(m_n)_{n\in\mathbb{N}}$ *is strongly universally consistent.*

Remark 1 Assumption (10) is satisfied, if (8) and (9) are sharpened to $h_n \downarrow 0$ and $n \cdot h_n^d \uparrow \infty$ $(n \to \infty)$. Indeed, if $h_n \downarrow 0$ holds, then (10) is equivalent to

$$\limsup_{n\to\infty} \frac{h_n}{h_{2\cdot n}} < \infty,$$

and in case that this condition does not hold we can find a subsequence $(n_k)_k$ of $(n)_n$ such that

$$\frac{2 \cdot n_k \cdot h_{2\cdot n_k}^d}{n_k \cdot h_{n_k}^d} = 2 \cdot \left(\frac{h_{2\cdot n_k}}{h_{n_k}}\right)^d \to 0 \quad (n \to \infty),$$

which implies that $n \cdot h_n^d \uparrow \infty$ $(n \to \infty)$ does not hold.

Theorem 3 *For* $n \in \mathbb{N}$ *let* $\mathcal{P}_n = \{A_{n,j} : j\}$ *be a partition of* $\mathbb{R}^d$*. Assume that for all* $x \in \mathbb{R}^d$

$$diam(A_{\mathcal{P}_n}(x)) \to 0 \quad (n \to \infty) \tag{11}$$

and that for each sphere S centered at the origin

$$\frac{1}{n} |\{A \in \mathcal{P}_n : A \cap S \neq \emptyset\}| \to 0 \quad (n \to \infty). \tag{12}$$

Assume furthermore that $(\mathcal{P}_n)_n$ *is nested in the sense that each set in* $\mathcal{P}_n$ *is the union of finitely many sets in* $\mathcal{P}_{n+1}$*, and that there exists an* $L \in \mathbb{N}$ *with the property that for each* $n \in \mathbb{N}$ *each set of* $\mathcal{P}_n$ *is the union of at most L sets in* $\mathcal{P}_{2\cdot n}$*. Define the partitioning regression estimate corresponding to* $\mathcal{P}_n$ *by*

$$m_n(x) = \frac{\sum_{i=1}^n I_{A_{\mathcal{P}_n}(x)}(X_i) \cdot Y_i}{\sum_{j=1}^n I_{A_{\mathcal{P}_n}(x)}(X_j)}.$$

Then $(m_n)_{n\in\mathbb{N}}$ *is strongly universally consistent.*

3 Proofs

3.1 Proof of Theorem 1

Lemma 23.3 in Györfi et al. (2002), which is due to Györfi (1991), with its proof together with the conditions $(A1)$ and $(A5)$, implies that it suffices to show the assertion in case that we have $|Y| \leq L$ a.s. for some $L > 0$. So from now on we assume that this condition holds.

We have

$$\int |m_n(x) - m(x)|^2 \, \mu(\mathrm{d}x)$$

$$
\begin{aligned}
&\leq 2L \cdot \int |m_n(x) - m(x)| \, \mu(\mathrm{d}x) \\
&\leq 2L \cdot \int \left| \sum_{i=1}^{n} W_{n,i}(x) \cdot (Y_i - m(X_i)) \right| \mu(\mathrm{d}x) \\
&\quad + 2L \cdot \int \left| \sum_{i=1}^{n} W_{n,i}(x) \cdot (m(X_i) - m(x)) \right| \mu(\mathrm{d}x) \\
&\quad + 2L \cdot \int \left| \left(\sum_{i=1}^{n} W_{n,i}(x) - 1 \right) \cdot m(x) \right| \mu(\mathrm{d}x) \\
&=: 2L \cdot J_n + 2L \cdot I_n + 2L \cdot M_n.
\end{aligned}
$$

Assumption $(A4)$ implies

$$M_n \to 0 \quad a.s.$$

Hence, it suffices to show

$$I_n \to 0 \quad a.s. \tag{13}$$

and

$$J_n \to 0 \quad a.s. \tag{14}$$

In order to prove (13), let $\epsilon > 0$ be arbitrary and choose a uniformly continuous function $\bar{m} : \mathbb{R}^d \to \mathbb{R}$ with bounded support such that $\int |\bar{m}(x) - m(x)| \mu(\mathrm{d}x) < \epsilon$ (cf., e.g., Györfi et al. 2002, Theorem A.1). By assumption $(A5)$ we get

$$
\begin{aligned}
I_n &\leq \int \left| \sum_{i=1}^{n} W_{n,i}(x) \cdot (m(X_i) - \bar{m}(X_i)) \right| \mu(\mathrm{d}x) \\
&\quad + \int \left| \sum_{i=1}^{n} W_{n,i}(x) \cdot (\bar{m}(X_i) - \bar{m}(x)) \right| \mu(\mathrm{d}x) \\
&\quad + \int \left| \sum_{i=1}^{n} W_{n,i}(x) \cdot (\bar{m}(x) - m(x)) \right| \mu(\mathrm{d}x) \\
&\leq \int \sum_{i=1}^{n} |W_{n,i}(x)| \cdot |m(X_i) - \bar{m}(X_i)| \, \mu(\mathrm{d}x) \\
&\quad + \int \sum_{i=1}^{n} |W_{n,i}(x)| \cdot |\bar{m}(X_i) - \bar{m}(x)| \, \mu(\mathrm{d}x) \\
&\quad + D \cdot \int |\bar{m}(x) - m(x))| \, \mu(\mathrm{d}x).
\end{aligned}
$$

Application of assumption $(A1)$ and the choice of $\bar{m}$ yields

$$\limsup_{n\to\infty} I_n \le c\cdot\epsilon + \limsup_{n\to\infty} \int \sum_{i=1}^{n} |W_{n,i}(x)|\cdot|\bar{m}(X_i)-\bar{m}(x)|\,\mu(\mathrm{d}x) + D\cdot\epsilon \quad a.s.$$

Because of $(A5)$, for arbitrary $\delta > 0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \int \sum_{i=1}^{n} |W_{n,i}(x)|\cdot|\bar{m}(X_i)-\bar{m}(x)|\,\mu(\mathrm{d}x)\\
&\quad\le \limsup_{n\to\infty}\Bigg[\int \sum_{i=1}^{n} |W_{n,i}(x)|\cdot|\bar{m}(X_i)-\bar{m}(x)|\cdot I_{\{\|X_i-x\|>\delta\}}\mu(\mathrm{d}x)\\
&\qquad+\int \sum_{i=1}^{n} |W_{n,i}(x)|\cdot|\bar{m}(X_i)-\bar{m}(x)|\cdot I_{\{\|X_i-x\|\le\delta\}}\mu(\mathrm{d}x)\Bigg]\\
&\quad\le 2\cdot\sup_{x\in\mathbb{R}^d}|\bar{m}(x)|\cdot\limsup_{n\to\infty}\int \sum_{i=1}^{n} |W_{n,i}(x)|\cdot I_{\{\|X_i-x\|>\delta\}}\mu(\mathrm{d}x)\\
&\qquad+D\cdot\left(\sup_{x,z\in\mathbb{R}^d,\|x-z\|\le\delta}|\bar{m}(x)-\bar{m}(z)|\right).
\end{aligned}$$

The first term on the right-hand side above is zero with probability one because of assumption $(A2)$. The second term will become arbitrarily small for small δ because of the smoothness of $\bar{m}$. From this and the above inequality, we get (13).

Now (14) will be shown. For an arbitrary $\epsilon > 0$ choose $c_3 > 0$ such that $c_3 < \epsilon^2/(2L^2)$. With the notation

$$B_n := \left\{\sum_{i=1}^{n}\left(\int |W_{n,i}(x)|\mu(\mathrm{d}x)\right)^2 \le c_3/\log(n)\right\}$$

we have

$$\limsup_{n\to\infty} J_n = \limsup_{n\to\infty}\left[J_n\cdot I_{B_n^c} + J_n\cdot I_{B_n}\right] \le 0 + \limsup_{n\to\infty} J_n\cdot I_{B_n} \quad a.s.$$

because of $(A3)$. It remains to show

$$\limsup_{n\to\infty} J_n\cdot I_{B_n} \le \epsilon \quad a.s.$$

Because of the lemma of Borel–Cantelli, it suffices to show

$$\sum_{n=1}^{\infty}\mathbf{P}\left(J_n\cdot I_{B_n} > \epsilon\right) < \infty.$$

Therefore, set

$$F(y_1, \ldots, y_n) = \int \left| \sum_{i=1}^{n} (y_i - m(X_i)) \cdot W_{n,i}(x) \right| \mu(\mathrm{d}x).$$

Then for every $i \in \{1, \ldots, n\}$

$$\left| F(y_1, \ldots, y_n) - F(y_1, \ldots, y_{i-1}, y_i', y_{i+1}, \ldots, y_n) \right| \leq \left| y_i - y_i' \right| \int |W_{n,i}(x)| \mu(\mathrm{d}x).$$

Using the inequality of McDiarmid (cf., e.g., McDiarmid 1989 or Theorem A.2 in Györfi et al. 2002), we get

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbf{P}\left(J_n \cdot I_{B_n} > \epsilon\right) \\
&\quad = \sum_{n=1}^{\infty} \mathbf{E}\left\{\mathbf{P}\left(J_n > \epsilon \mid X_1, \ldots, X_n\right) \cdot I_{B_n}\right\} \\
&\quad \leq \sum_{n=1}^{\infty} \mathbf{E}\left\{2 \cdot \exp\left(\frac{-2\epsilon^2}{4L^2 \cdot \sum_{i=1}^{n} \left(\int |W_{n,i}(x)| \mu(\mathrm{d}x)\right)^2}\right) \cdot I_{B_n}\right\} \\
&\quad \leq 2 \sum_{n=1}^{\infty} \mathbf{E}\left\{\exp\left(\frac{-\epsilon^2 \log(n)}{2L^2 \cdot c_3}\right)\right\} \\
&\quad < \infty.
\end{aligned}$$

The proof is complete. □

Remark 2 Assumption $(A3)$ in Theorem 1 follows from

$$(A3') \qquad \sum_{n=1}^{\infty} (\log(n))^2 \cdot n \cdot \sum_{i=1}^{n} \mathbf{E}\left\{\left(\int |W_{n,i}(x)| \mu(\mathrm{d}x)\right)^4\right\} < \infty$$

by using the lemma of Borel–Cantelli together with the inequality of Markov and Jensen's inequality. On the other hand, in Theorem 1 the assumption $(A3)$ may be replaced by

$$\sum_{n=1}^{\infty} n \cdot \sum_{i=1}^{n} \mathbf{E}\left\{\left(\int |W_{n,i}(x)| \mu(\mathrm{d}x)\right)^4\right\} < \infty.$$

To show this, in the proof of (14) one avoids the introduction of B_n and notices

$$2 \cdot \exp\left(\frac{-\epsilon^2}{2L^2 \cdot \sum_{i=1}^{n} \left(\int |W_{n,i}(x)|\mu(\mathrm{d}x)\right)^2}\right) \leq \frac{8 \cdot L^4}{\epsilon^4}\left(\sum_{i=1}^{n}\left(\int |W_{n,i}(x)|\mu(\mathrm{d}x)\right)^2\right)^2$$

$$\leq \frac{8 \cdot L^4}{\epsilon^4} \cdot n \cdot \sum_{i=1}^{n}\left(\int |W_{n,i}(x)|\mu(\mathrm{d}x)\right)^4$$

because of $z^2 \cdot e^{-z} \leq 1$ $(z \in \mathbb{R}_+)$.

Remark 3 As mentioned above, choice of the L_2 metric is induced by the minimum property of the regression function m. An obvious modification of the proof of Theorem 1 yields strong consistency in the L_p norm $(1 \leq p < \infty)$ under the assumption $\mathbf{E}\{|Y|^p\} < \infty$.

3.2 Proof of Theorem 2

The crucial part in the proof of Theorem 2 is the verification of "Stone's technical condition" (A1). It is done by use of Etemadi's "thinning" argument and the covering Lemma 4.

Lemma 1 *Assume that the assumptions of Theorem 2 hold and set*

$$W_{n,i}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_n}\right)}.$$

Then $W_{n,i}$ satisfies the assumption (A1) *from Theorem 1.*

In the proof of Lemma 1 we will need the following auxiliary results (Lemmas 2, 3, 4 and 5).

Lemma 2 *Let $n \in \mathbb{N}$ and $p \in (0, 1]$. Then we have:*

(a)

$$\sum_{k=0}^{n} \frac{1}{1+k}\binom{n}{k} p^k (1-p)^{n-k} = \frac{1-(1-p)^{n+1}}{(n+1) \cdot p}.$$

(b)

$$\sum_{k=0}^{n} \frac{1}{(1+k)^2}\binom{n}{k} p^k (1-p)^{n-k} \leq \frac{2}{(n+1)^2 \cdot p^2}.$$

(c)

$$\sum_{k=0}^{n} \frac{1}{(1+k)^4} \binom{n}{k} p^k (1-p)^{n-k} \leq \frac{24}{(n+1)^4 \cdot p^4}.$$

Proof (a) We have

$$\begin{aligned}\sum_{k=0}^{n} \frac{1}{1+k} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{1}{(n+1) \cdot p} \sum_{k=0}^{n} \binom{n+1}{k+1} p^{k+1} (1-p)^{n+1-(k+1)} \\ &= \frac{1-(1-p)^{n+1}}{(n+1) \cdot p}.\end{aligned}$$

(b) We have

$$\begin{aligned}&\sum_{k=0}^{n} \frac{1}{(1+k)^2} \binom{n}{k} p^k (1-p)^{n-k} \\ &\quad \leq \sum_{k=0}^{n} \frac{2}{(k+2) \cdot (k+1)} \binom{n}{k} p^k (1-p)^{n-k} \\ &\quad = \frac{2}{(n+1) \cdot (n+2) \cdot p^2} \cdot \sum_{k=0}^{n} \binom{n+2}{k+2} p^{k+2} (1-p)^{n+2-(k+2)} \\ &\quad \leq \frac{2}{(n+1)^2 \cdot p^2} \cdot 1.\end{aligned}$$

(c) We have

$$\begin{aligned}&\sum_{k=0}^{n} \frac{1}{(1+k)^4} \binom{n}{k} p^k (1-p)^{n-k} \\ &\leq \sum_{k=0}^{n} \frac{24}{(k+4) \cdot (k+3) \cdot (k+2) \cdot (k+1)} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{24}{(n+1) \cdot (n+2) \cdot (n+3) \cdot (n+4) \cdot p^4} \cdot \sum_{k=0}^{n} \binom{n+4}{k+4} p^{k+4} (1-p)^{n+4-(k+4)} \\ &\leq \frac{24}{(n+1)^4 \cdot p^4} \cdot 1.\end{aligned}$$

□

Lemma 3 *Let* $n \in \mathbb{N}$ *and* $p_1, p_2 > 0$ *such that* $p_1 + p_2 \leq 1$. *Then*

$$\sum_{\substack{k_1,k_2 \in \{0,\ldots,n\} \\ k_1+k_2 \leq n}} \frac{1}{(1+k_1) \cdot (1+k_2)} \cdot \frac{n!}{k_1! k_2! (n-(k_1+k_2))!} \cdot p_1^{k_1} \cdot p_2^{k_2}$$

$$\cdot(1-(p_1+p_2))^{n-(k_1+k_2)}$$
$$= \frac{1}{(n+1)\cdot(n+2)\cdot p_1\cdot p_2}\cdot\Big(1-(1-p_1)^{n+2}-(1-p_2)^{n+2}$$
$$+(1-(p_1+p_2))^{n+2}\Big).$$

Proof We have

$$\sum_{\substack{k_1,k_2\in\{0,\dots,n\}\\ k_1+k_2\le n}} \frac{1}{(1+k_1)\cdot(1+k_2)}\cdot\frac{n!}{k_1!k_2!(n-(k_1+k_2))!}\cdot p_1^{k_1}\cdot p_2^{k_2}$$
$$\cdot(1-(p_1+p_2))^{n-(k_1+k_2)}$$
$$= \frac{1}{(n+1)\cdot(n+2)\cdot p_1\cdot p_2}$$
$$\cdot\sum_{\substack{k_1,k_2\in\{0,\dots,n\}\\ k_1+k_2\le n}} \frac{(n+2)!}{(k_1+1)!(k_2+1)!(n+2-(k_1+1+k_2+1))!}$$
$$\cdot p_1^{k_1+1}\cdot p_2^{k_2+1}\cdot(1-(p_1+p_2))^{n+2-(k_1+1+k_2+1)}$$
$$= \frac{1}{(n+1)\cdot(n+2)\cdot p_1\cdot p_2}\cdot\sum_{\substack{k_1,k_2\in\{1,\dots,n+1\}\\ k_1+k_2\le n+2}} \frac{(n+2)!}{k_1!k_2!(n+2-(k_1+k_2))!}$$
$$\cdot p_1^{k_1}\cdot p_2^{k_2}\cdot(1-(p_1+p_2))^{n+2-(k_1+k_2)}.$$

From this and

$$\sum_{\substack{k_1,k_2\in\{1,\dots,n+1\}\\ k_1+k_2\le n+2}} \frac{(n+2)!}{k_1!k_2!(n+2-(k_1+k_2))!}\cdot p_1^{k_1}\cdot p_2^{k_2}\cdot(1-(p_1+p_2))^{n+2-(k_1+k_2)}$$
$$= \sum_{\substack{k_1,k_2\in\{0,\dots,n+2\}\\ k_1+k_2\le n+2}} \frac{(n+2)!}{k_1!k_2!(n+2-(k_1+k_2))!}\cdot p_1^{k_1}\cdot p_2^{k_2}$$
$$\cdot(1-(p_1+p_2))^{n+2-(k_1+k_2)}$$
$$-\sum_{k_1=0}^{n+2}\binom{n+2}{k_1}p_1^{k_1}(1-(p_1+p_2))^{n+2-k_1}$$
$$-\sum_{k_2=0}^{n+2}\binom{n+2}{k_2}p_2^{k_2}(1-(p_1+p_2))^{n+2-k_2}$$
$$+(1-(p_1+p_2))^{n+2}$$
$$= 1-(p_1+(1-(p_1+p_2)))^{n+2}-(p_2+(1-(p_1+p_2)))^{n+2}$$
$$+(1-(p_1+p_2))^{n+2}$$

we get the assertion. □

Lemma 4 (a) *Let $\mathcal{P}$, $\mathcal{P}^*$ be partitions of $\mathbb{R}^d$ such that there exists an $L \in \mathbb{N}$ with the property that each set of $\mathcal{P}^*$ is contained in the union of at most L sets in $\mathcal{P}$. Then for all $t \in \mathbb{R}^d$ one has:*

$$\int_{\mathbb{R}^d} \frac{I_{A_{\mathcal{P}^*}(t)}(x)}{\mu(A_{\mathcal{P}}(x))}\, \mu(dx) \leq L.$$

(b) *Let $0 < r_1 \leq r_2$. Then there exists an $L \in \mathbb{N}$ such that for all $n \in \mathbb{N}$ and all $t \in \mathbb{R}$ one has*

$$\int_{\mathbb{R}^d} \frac{I_{t+h_n \cdot S_{r_2}}(x)}{\mu(x + h_n \cdot S_{r_1})}\, \mu(dx) \leq L.$$

Proof (a) Choose $t_1, \dots t_L$ such that

$$A_{\mathcal{P}^*}(t) \subseteq \cup_{l=1}^{L} A_{\mathcal{P}}(t_l).$$

Then

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{I_{A_{\mathcal{P}^*}(t)}(x)}{\mu(A_{\mathcal{P}}(x))}\, \mu(\mathrm{d}x) &\leq \sum_{l=1}^{L} \int_{\mathbb{R}^d} \frac{I_{A_{\mathcal{P}}(t_l)}(x)}{\mu(A_{\mathcal{P}}(x))}\, \mu(\mathrm{d}x) \\ &= \sum_{l=1}^{L} \int_{\mathbb{R}^d} \frac{I_{A_{\mathcal{P}}(t_l)}(x)}{\mu(A_{\mathcal{P}}(t_l))}\, \mu(\mathrm{d}x) \leq L,\end{aligned}$$

since $x \in A_{\mathcal{P}}(t_l)$ implies $A_{\mathcal{P}}(x) = A_{\mathcal{P}}(t_l)$.
(b) Choose $L \in \mathbb{N}$ and cubes $A_1, \dots, A_L \subseteq \mathbb{R}^d$ of side length $r_1/\sqrt{d}$ such that

$$L \leq \left(\frac{2 \cdot r_2}{r_1/\sqrt{d}} + 2\right)^d \quad \text{and} \quad S_{r_2} \subseteq \cup_{l=1}^{L} A_l.$$

Then

$$\begin{aligned}&\int \frac{I_{t+h_n \cdot S_{r_2}}(x)}{\mu\left(x + h_n \cdot S_{r_1}\right)}\, \mu(\mathrm{d}x) \\ &\leq \sum_{l=1}^{L} \int \frac{I_{t+h_n \cdot A_l}(x)}{\mu\left(x + h_n \cdot S_{r_1}\right)}\, \mu(\mathrm{d}x) \\ &\leq \sum_{l=1}^{L} \int \frac{I_{t+h_n \cdot A_l}(x)}{\mu\left(t + h_n \cdot A_l\right)}\, \mu(\mathrm{d}x) \\ &\leq L,\end{aligned}$$

where the second inequality followed from $x + h_n \cdot S_{r_1} \supseteq t + h_n \cdot A_l$ for $x \in t + h_n \cdot A_l$. □

Lemma 5 *Let* $0 < r_1 \leq r_2$. *Then constants* $c_4, c_5 > 0$ *exist such that for each* $i \in \{1, \dots, n\}$

$$\mathbf{E}\left\{\left(\int \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}\mu(dx)\right)^2\right\} \leq \frac{c_4}{n^2}$$

and

$$\mathbf{E}\left\{\left(\int \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}\mu(dx)\right)^4\right\} \leq \frac{c_5}{n^4}.$$

Proof Because of the Cauchy–Schwarz inequality, it suffices to show the second inequality. By using the independence of the data, by applying twice the Cauchy–Schwarz inequality and by Lemma 2 c), we get for $i \in \{1, \dots, n\}$:

$$\begin{aligned}
&\mathbf{E}\left\{\left(\int \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}\,\mu(\mathrm{d}x)\right)^4\right\}\\
&= \mathbf{E}\Bigg\{\int\int\int\int \prod_{l=1}^{4} \frac{I_{S_{r_2}}\left(\frac{x^{(l)}-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x^{(l)}-X_j}{h_n}\right)}\\
&\quad \mu(\mathrm{d}x^{(1)})\,\mu(\mathrm{d}x^{(2)})\,\mu(\mathrm{d}x^{(3)})\,\mu(\mathrm{d}x^{(4)})\Bigg\}\\
&= \int\int\int\int \mathbf{E}\left\{\prod_{l=1}^{4} I_{S_{r_2}}\left(\frac{x^{(l)}-X_i}{h_n}\right)\right\}\cdot\\
&\quad \mathbf{E}\left\{\prod_{l=1}^{4} \frac{1}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x^{(l)}-X_j}{h_n}\right)}\right\}\\
&\quad \mu(\mathrm{d}x^{(1)})\,\mu(\mathrm{d}x^{(2)})\,\mu(\mathrm{d}x^{(3)})\,\mu(\mathrm{d}x^{(4)})\\
&\leq \int\int\int\int \mathbf{E}\left\{\prod_{l=1}^{4} I_{S_{r_2}}\left(\frac{x^{(l)}-X_i}{h_n}\right)\right\}\\
&\quad \cdot\prod_{l=1}^{4}\left(\mathbf{E}\left\{\frac{1}{\left(1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x^{(l)}-X_j}{h_n}\right)\right)^4}\right\}\right)^{\frac{1}{4}}
\end{aligned}$$

$$
\begin{aligned}
&\mu(\mathrm{d}x^{(1)})\,\mu(\mathrm{d}x^{(2)})\,\mu(\mathrm{d}x^{(3)})\,\mu(\mathrm{d}x^{(4)})\\
&\le \int\int\int\int \mathbf{E}\left\{\prod_{l=1}^{4} I_{S_{r_2}}\left(\frac{x^{(l)}-X_i}{h_n}\right)\right\}\cdot\prod_{l=1}^{4}\frac{24^{1/4}}{n\cdot\mu\left(x^{(l)}+h_n\cdot S_{r_1}\right)}\\
&\quad\mu(\mathrm{d}x^{(1)})\,\mu(\mathrm{d}x^{(2)})\,\mu(\mathrm{d}x^{(3)})\,\mu(\mathrm{d}x^{(4)})\\
&= \frac{24}{n^4}\cdot\mathbf{E}\left\{\prod_{l=1}^{4}\int\frac{I_{S_{r_2}}\left(\frac{x^{(l)}-X_i}{h_n}\right)}{\mu\left(x^{(l)}+h_n\cdot S_{r_1}\right)}\mu(\mathrm{d}x^{(l)})\right\}\\
&= \frac{24}{n^4}\cdot\mathbf{E}\left\{\left(\int\frac{I_{X_i+h_n\cdot S_{r_2}}(x)}{\mu\left(x+h_n\cdot S_{r_1}\right)}\mu(\mathrm{d}x)\right)^4\right\}.
\end{aligned}
$$

Application of Lemma 4b yields the assertion. □

Proof of Lemma 1. By rescaling the kernel, if necessary, we can assume w.l.o.g. that we have $c_2 \le 1$. Let Y be such that $Y \ge 0$ a.s. and $\mathbf{E}Y < \infty$. For $a \in [0,1]$ and $b \ge 0$ we have

$$
\frac{a}{a+b} \le \frac{1}{1+\frac{b}{a}} \le \frac{1}{1+b}
$$

(noticing $0/0 = 0$), which implies for $i \in \{1,\ldots,n\}$

$$
\begin{aligned}
\frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^{n}K\left(\frac{x-X_j}{h_n}\right)} &= \frac{K\left(\frac{x-X_i}{h_n}\right)}{K\left(\frac{x-X_i}{h_n}\right)+\sum_{j\in\{1,\ldots,n\}\setminus\{i\}}K\left(\frac{x-X_j}{h_n}\right)}\cdot I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)\\
&\le \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\ldots,n\}\setminus\{i\}}K\left(\frac{x-X_j}{h_n}\right)}\\
&\le \frac{1}{c_1}\cdot\frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\ldots,n\}\setminus\{i\}}I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}.
\end{aligned}
\tag{15}
$$

Consequently, we have

$$
\sum_{i=1}^{n}W_{n,i}(x)\cdot Y_i \le \frac{1}{c_1}\cdot\sum_{i=1}^{n}\frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)\cdot Y_i}{1+\sum_{j\in\{1,\ldots,n\}\setminus\{i\}}I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}
$$

and it suffices to show for some $c_6 > 0$

$$
\limsup_{n\to\infty}\int\sum_{i=1}^{n}\frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)\cdot Y_i}{1+\sum_{j\in\{1,\ldots,n\}\setminus\{i\}}I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}\,\mu(\mathrm{d}x) \le c_6\cdot\mathbf{E}Y. \tag{16}
$$

Set $Y_i^* = Y_i \cdot I_{\{Y_i \le i\}}$. We show next that (16) follows from

$$\limsup_{n\to\infty} \int \sum_{i=1}^{n} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right) \cdot Y_i^*}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)} \mu(dx) \le c_6 \cdot \mathbf{E}Y. \tag{17}$$

To do this, we observe that

$$\sum_{n=1}^{\infty} \mathbf{P}\{Y_n^* \ne Y_n\} = \sum_{n=1}^{\infty} \mathbf{P}\{Y_n > n\} \le \int_0^{\infty} \mathbf{P}\{Y > t\}\, dt = \mathbf{E}Y < \infty$$

implies that we have with probability one that $Y_n^* \ne Y_n$ holds at most for finitely many n. For each $i \in \mathbb{N}$, one has

$$\int \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)} \mu(dx) \to 0 \quad a.s., \tag{18}$$

which follows via the lemma of Borel–Cantelli and Markov inequality from

$$\sum_{n=i}^{\infty} \mathbf{E}\left\{\left(\int \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)} \mu(dx)\right)^2\right\} \le c_4 \cdot \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty,$$

where the first inequality is a consequence of Lemma 5.

From (18) together with $Y_n - Y_n^* \ne 0$ only for finitely many n almost surely and (17), we can conclude

$$\begin{aligned}
&\limsup_{n\to\infty} \int \sum_{i=1}^{n} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right) \cdot Y_i}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)} \mu(dx) \\
&\le \limsup_{n\to\infty} \int \sum_{i=1}^{n} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right) \cdot Y_i^*}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)} \mu(dx) \\
&\quad + \limsup_{n\to\infty} \int \sum_{i=1}^{n} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right) \cdot (Y_i - Y_i^*)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)} \mu(dx) \\
&\le c_6 \cdot \mathbf{E}Y + 0 \quad a.s.
\end{aligned}$$

In the sequel we show (17). Set $n_k = 2^k$ ($k \in \mathbb{N}$). Then we have for any $n \in \{n_k, n_k + 1, \dots, n_{k+1}\}$

$$\sum_{i=1}^{n} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right) \cdot Y_i^*}{1 + \sum_{j \in \{1,\dots,n\} \setminus \{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}$$
$$\leq \sum_{i=1}^{n_{k+1}} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{h_n}\right) \cdot Y_i^*}{1 + \sum_{j \in \{1,\dots,n_k\} \setminus \{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{h_n}\right)}$$
$$\leq \sum_{i=1}^{n_{k+1}} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{\max\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\}}\right) \cdot Y_i^*}{1 + \sum_{j \in \{1,\dots,n_k\} \setminus \{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{\min\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\}}\right)}.$$

Hence, in order to show (17) it suffices to show

$$\limsup_{k \to \infty} \int \sum_{i=1}^{n_{k+1}} \frac{I_{S_{r_2}}\left(\frac{x-X_i}{\max\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\}}\right) \cdot Y_i^*}{1 + \sum_{j \in \{1,\dots,n_k\} \setminus \{i\}} I_{S_{r_1}}\left(\frac{x-X_j}{\min\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\}}\right)} \mu(\mathrm{d}x) \leq c_6 \cdot \mathbf{E}Y,$$

which is equivalent to

$$\limsup_{k \to \infty} \int \sum_{i=1}^{n_{k+1}} \frac{I_{x+S_{\max\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\} \cdot r_2}}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\dots,n_k\} \setminus \{i\}} I_{x+S_{\min\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\} \cdot r_1}}(X_j)} \mu(\mathrm{d}x) \leq c_6 \cdot \mathbf{E}Y. \tag{19}$$

Set $r_1' = \min\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\} \cdot r_1$ and $r_2' = \max\{h_{n_k}, h_{n_k+1}, \dots, h_{n_{k+1}}\} \cdot r_2$. Let $\mathcal{P}_{n_{k+1}}$ be a partition of $\mathbb{R}^d$ into cubes of sidelength $r_1'/\sqrt{d}$. Then we have for any $x \in \mathbb{R}^d$:

$$A_{\mathcal{P}_{n_{k+1}}}(x) \subseteq x + S_{r_1'}.$$

Let $\mathcal{P}_{n_{k+1}}^*$ be a partition of $\mathbb{R}^d$ into cubes of sidelength $2 \cdot r_2'$ and choose shifted versions $\mathcal{P}^{*(1)}, \dots, \mathcal{P}^{*(2^d)}$ of $\mathcal{P}_{n_{k+1}}^*$ such that for any $x \in \mathbb{R}^d$ there exist cells $A^{(1)} \in \mathcal{P}^{*(1)}, \dots, A^{(2^d)} \in \mathcal{P}^{*(2^d)}$ with

$$x + S_{r_2'} \subseteq \cup_{l=1}^{2^d} A_{\mathcal{P}^{*(l)}}(x).$$

(Here the shifted partitions can be chosen by choosing first a subset of the d components of x and by shifting then all sets in each coordinate contained in the subset by r_2' away from zero.) Consequently, we have

$$\sum_{i=1}^{n_{k+1}} \frac{I_{x+S_{r_2'}}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\dots,n_k\} \setminus \{i\}} I_{x+S_{r_1'}}(X_j)}$$

$$\leq \sum_{i=1}^{n_{k+1}} \sum_{l=1}^{2^d} \frac{I_{A_{\mathcal{P}^*(l)}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}$$
$$= \sum_{l=1}^{2^d} \sum_{i=1}^{n_{k+1}} \frac{I_{A_{\mathcal{P}^*(l)}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}.$$

This shows that it suffices to show for some $c_7 > 0$

$$\limsup_{k\to\infty} \int \sum_{i=1}^{n_{k+1}} \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)} \, \mu(\mathrm{d}x) \leq c_7 \cdot \mathbf{E}Y, \tag{20}$$

where $\mathcal{P}_{n_{k+1}}$ and $\mathcal{P}^*_{n_{k+1}}$ are partitions of $\mathbb{R}^d$ such that for each set $A \in \mathcal{P}^*_{n_{k+1}}$ there exists $L \in \mathbb{N}$ and sets $A_1, \ldots, A_L \in \mathcal{P}_{n_{k+1}}$ with the properties

$$A \subseteq \cup_{l=1}^{L} A_l \quad \text{and} \quad L \leq \left(\frac{2 \cdot r_2'}{r_1'/\sqrt{d}} + 2\right)^d \leq c_8 \cdot \left(\frac{\max\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}{\min\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}\right)^d.$$

By using the independence of the data and by Lemmas 2a and 4a we have

$$\mathbf{E} \int \sum_{i=1}^{n_{k+1}} \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)} \, \mu(\mathrm{d}x)$$
$$\leq n_{k+1} \cdot \int \frac{\mathbf{E}\left\{Y_1 \cdot I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_1)\right\}}{n_k \cdot \mu\left(A_{\mathcal{P}_{n_{k+1}}}(x)\right)} \, \mu(\mathrm{d}x)$$
$$= \frac{n_{k+1}}{n_k} \cdot \int m(t) \cdot \int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(t)}(x)}{\mu\left(A_{\mathcal{P}_{n_{k+1}}}(x)\right)} \, \mu(\mathrm{d}x) \, \mu(dt)$$
$$\leq 2 \cdot c_8 \cdot \left(\frac{\max\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}{\min\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}\right)^d \int m(t) \, \mu(dt)$$
$$= 2 \cdot c_8 \cdot \left(\frac{\max\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}{\min\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}\right)^d \cdot \mathbf{E}Y.$$

Hence, because of (10), it suffices to show

$$\sum_{i=1}^{n_{k+1}} \int \left(\frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)} \right.$$
$$\left. - \mathbf{E}\left\{ \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)} \right\} \right) \mu(\mathrm{d}x) \to 0 \quad a.s.$$

An application of the lemma of Borel–Cantelli and the inequality of Chebyshev yields that this in turn follows from

$$\sum_{k=1}^{\infty} \mathbf{V}\left\{\sum_{i=1}^{n_{k+1}} \int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right\} < \infty. \tag{21}$$

So in order to finish the proof of Lemma 1, it remains to show (21). This will be done in the sequel.

We have

$$\begin{aligned}
&\mathbf{V}\left\{\sum_{i=1}^{n_{k+1}} \int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right\}\\
&= \sum_{i=1}^{n_{k+1}} \mathbf{V}\left\{\int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i) \cdot Y_i^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right\}\\
&\quad+\sum_{\substack{l,r\in\{1,\ldots,n_{k+1}\}\\ l\neq r}} \left(\mathbf{E}\left\{\int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_l) \cdot Y_l^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right.\right.\\
&\quad\cdot\left.\int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_r) \cdot Y_r^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right\}\\
&\quad-\mathbf{E}\left\{\int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_l) \cdot Y_l^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right\}\\
&\quad\left.\cdot\mathbf{E}\left\{\int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_r) \cdot Y_r^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\,\mu(\mathrm{d}x)\right\}\right)\\
&=: \sum_{i=1}^{n_{k+1}} B_{i,k} + \sum_{\substack{l,r\in\{1,\ldots,n_{k+1}\}\\ l\neq r}} D_{k,l,r};
\end{aligned}$$

hence, it suffices to show

$$\sum_{k=1}^{\infty}\sum_{i=1}^{n_{k+1}} B_{i,k} < \infty \tag{22}$$

and

$$\sum_{k=1}^{\infty}\sum_{\substack{l,r\in\{1,\ldots,n_{k+1}\}\\ l\neq r}} D_{k,l,r} < \infty. \tag{23}$$

By the independence of the data, by the inequality of Cauchy–Schwarz and by an application of Lemmas 2b and 4a we get

$$\sum_{i=1}^{n_{k+1}} B_{i,k}$$
$$\leq \sum_{i=1}^{n_{k+1}} \mathbf{E}\left\{\left(\int \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i)\cdot Y_i^*}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}\mu(\mathrm{d}x)\right)^2\right\}$$
$$= \sum_{i=1}^{n_{k+1}} \int\int \mathbf{E}\Bigg\{(Y_i^*)^2 \cdot \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_i)}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}$$
$$\cdot \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X_i)}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}\left(X_j\right)}\Bigg\}\mu(dz)\,\mu(\mathrm{d}x)$$
$$= \sum_{i=1}^{n_{k+1}} \int\int\int \mathbf{E}\{(Y_i^*)^2|X_i=t\}\cdot \mathbf{E}\Bigg\{\frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(t)}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}\left(X_j\right)}$$
$$\cdot \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(t)}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}\left(X_j\right)}\Bigg\}\mu(dt)\,\mu(dz)\,\mu(\mathrm{d}x)$$
$$\leq \sum_{i=1}^{n_{k+1}} \int\int\int \mathbf{E}\{(Y_i^*)^2|X_i=t\}\cdot I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(t)\cdot I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(t)$$
$$\cdot\sqrt{\mathbf{E}\left\{\frac{1}{\left(1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(t)}\left(X_j\right)\right)^2}\right\}}$$
$$\cdot\sqrt{\mathbf{E}\left\{\frac{1}{\left(1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}\left(X_j\right)\right)^2}\right\}}\,\mu(dt)\,\mu(dz)\,\mu(\mathrm{d}x)$$
$$\leq \sum_{i=1}^{n_{k+1}} \int\int\int \mathbf{E}\{(Y_i^*)^2|X_i=t\}$$
$$\cdot 2\cdot \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(t)\cdot I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(t)}{n_k^2\cdot\mu(A_{\mathcal{P}_{n_{k+1}}}(x))\cdot\mu(A_{\mathcal{P}_{n_{k+1}}}(z))}\,\mu(dt)\,\mu(dz)\,\mu(\mathrm{d}x)$$
$$\leq \sum_{i=1}^{n_{k+1}} 2\cdot c_8^2\cdot\left(\frac{\max\{h_{n_k},h_{n_k+1},\ldots,h_{n_{k+1}}\}}{\min\{h_{n_k},h_{n_k+1},\ldots,h_{n_{k+1}}\}}\right)^{2\cdot d}\frac{\mathbf{E}\{(Y_i^*)^2\}}{n_k^2}.$$

This implies

$$\sum_{k=1}^{\infty}\sum_{i=1}^{n_{k+1}} B_{i,k} \leq c_9\cdot\sum_{k=1}^{\infty}\sum_{i=1}^{n_{k+1}}\frac{\mathbf{E}\{(Y_i^*)^2\}}{n_k^2}.$$

Since

$$\sum_{k=1}^{\infty}\sum_{i=1}^{n_{k+1}}\frac{\mathbf{E}\{(Y_i^*)^2\}}{n_k^2} \le 2\cdot\sum_{k=1}^{\infty}\frac{\mathbf{E}\{(Y_{n_{k+1}}^*)^2\}}{n_k}$$
$$= 4\cdot\sum_{k=1}^{\infty}\frac{\mathbf{E}\{(Y_{2^{k+1}}^*)^2\}}{2^{k+1}} = 4\cdot\mathbf{E}\left\{Y^2\cdot\sum_{k=1}^{\infty} I_{\{Y\le 2^{k+1}\}}\cdot\frac{1}{2^{k+1}}\right\}$$
$$= 4\cdot\mathbf{E}\left\{Y^2\cdot\sum_{k=\lceil\log_2 Y\rceil-1}^{\infty}\frac{1}{2^{k+1}}\right\} = 4\cdot\mathbf{E}\left\{Y^2\cdot\frac{1}{2^{\lceil\log_2 Y\rceil}}\right\}$$
$$\le 4\cdot\mathbf{E}Y < \infty,$$

this proves (22).

So it remains to prove (23). To do this, we observe that the independence of the data implies

$$\begin{aligned}
&D_{k,l,r}\\
&\le \int\int \mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_l)\cdot Y_l^*\right\}\cdot\mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X_r)\cdot Y_r^*\right\}\\
&\quad\cdot\left(\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\dots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}\right.\right.\\
&\quad\left.\cdot\frac{1}{1+\sum_{j\in\{1,\dots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)}\right\}\\
&\quad-\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\dots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}\right\}\\
&\quad\left.\cdot\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\dots,n_k\}\setminus\{r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)}\right\}\right)\mu(\mathrm{d}x)\mu(dz)\\
&\le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d\setminus A_{\mathcal{P}_{n_{k+1}}}(z)}\Big(\dots\Big)\mu(\mathrm{d}x)\,\mu(dz)\\
&\quad+\int_{\mathbb{R}^d}\int_{A_{\mathcal{P}_{n_{k+1}}}(z)}\Big(\dots\Big)\mu(\mathrm{d}x)\,\mu(dz)\\
&=: D^{(1)}_{k,l,r} + D^{(2)}_{k,l,r}.
\end{aligned}$$

Let $x, z \in \mathbb{R}^d$ and set $p_1 = \mu(A_{\mathcal{P}_{n_{k+1}}}(x))$ and $p_2 = \mu(A_{\mathcal{P}_{n_{k+1}}}(z))$. It is easy to see that in case $p_1 = 0$ or $p_2 = 0$ we have

$$\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\dots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}\cdot\frac{1}{1+\sum_{j\in\{1,\dots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)}\right\}$$

$$-\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}\right\}$$
$$\cdot\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)}\right\}$$
$$= 0;$$

hence, we assume in the sequel w.l.o.g. that we have $p_1 > 0$ and $p_2 > 0$.

In case $x \notin A_{\mathcal{P}_{n_{k+1}}}(z)$ we have $A_{\mathcal{P}_{n_{k+1}}}(x) \cap A_{\mathcal{P}_{n_{k+1}}}(z) = \emptyset$, and we can conclude by Lemmas 2a and 3

$$\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}\cdot\frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)}\right\}$$
$$-\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}\right\}$$
$$\cdot\mathbf{E}\left\{\frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)}\right\}$$
$$= \frac{1}{n_k\cdot(n_k-1)\cdot p_1\cdot p_2}\cdot\left(1-(1-p_1)^{n_k}-(1-p_2)^{n_k}+(1-(p_1+p_2))^{n_k}\right)$$
$$-\frac{1}{n_k\cdot p_1}\cdot(1-(1-p_1)^{n_k})\cdot\frac{1}{n_k\cdot p_2}\cdot(1-(1-p_2)^{n_k})$$
$$= \frac{n_k-n_k\cdot(1-p_1)^{n_k}-n_k\cdot(1-p_2)^{n_k}+n_k\cdot(1-(p_1+p_2))^{n_k}}{{n_k}^2\cdot(n_k-1)\cdot p_1\cdot p_2}$$
$$-\frac{(n_k-1)\cdot(1-(1-p_1)^{n_k}-(1-p_2)^{n_k}+(1-p_1)^{n_k}\cdot(1-p_2)^{n_k})}{{n_k}^2\cdot(n_k-1)\cdot p_1\cdot p_2}$$
$$= \frac{1-(1-p_1)^{n_k}-(1-p_2)^{n_k}+n_k\cdot(1-(p_1+p_2))^{n_k}-(n_k-1)\cdot(1-p_1)^{n_k}\cdot(1-p_2)^{n_k}}{{n_k}^2\cdot(n_k-1)\cdot p_1\cdot p_2}$$
$$\le \frac{1-(1-p_1)^{n_k}-(1-p_2)^{n_k}+(1-(p_1+p_2))^{n_k}}{{n_k}^2\cdot(n_k-1)\cdot p_1\cdot p_2}$$
$$\le \frac{1}{{n_k}^2\cdot(n_k-1)\cdot p_1\cdot p_2},$$

where the first inequality followed from $(1-(p_1+p_2)) \le (1-p_1)\cdot(1-p_2)$. Consequently, application of Lemma 4a yields

$$D^{(1)}_{l,k,r}$$
$$\le \int_{\mathbb{R}^d}\int_{\mathbb{R}^d\setminus A_{\mathcal{P}_{n_{k+1}}}(z)}\left(\mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(x)}(X_l)\cdot Y^*_l\right\}\cdot\mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X_r)\cdot Y^*_r\right\}\right.$$
$$\left.\cdot\frac{1}{n_k^2\cdot(n_k-1)\cdot\mu(A_{\mathcal{P}_{n_{k+1}}}(x))\cdot\mu(A_{\mathcal{P}_{n_{k+1}}}(z))}\right)\mu(\mathrm{d}x)\,\mu(dz)$$
$$\le \frac{1}{n_k^2\cdot(n_k-1)}\cdot\int\int\int\int \mathbf{E}\{Y|X=t\}\cdot\mathbf{E}\{Y|X=s\}$$

$$\cdot \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}(x)}}(t)}{\mu(A_{\mathcal{P}_{n_{k+1}}}(x))} \cdot \frac{I_{A_{\mathcal{P}^*_{n_{k+1}}(z)}}(s)}{\mu(A_{\mathcal{P}_{n_{k+1}}}(z))} \cdot \mu(dt)\,\mu(ds)\,\mu(\mathrm{d}x)\,\mu(dz)$$

$$\leq \frac{1}{n_k^2 \cdot (n_k - 1)} \cdot c_8^2 \cdot \left(\frac{\max\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}{\min\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}} \right)^{2d} \cdot (\mathbf{E}Y)^2.$$

In case $x \in A_{\mathcal{P}_{n_{k+1}}}(z)$ we have $A_{\mathcal{P}_{n_{k+1}}}(x) = A_{\mathcal{P}_{n_{k+1}}}(z)$. Using this,

$$\begin{aligned} \frac{1}{(1+i)^2} &= \frac{1}{(i+1)\cdot(i+2)} + \frac{1}{(i+1)^2\cdot(i+2)} \\ &\leq \frac{1}{(i+1)\cdot(i+2)} + \frac{3}{(i+1)\cdot(i+2)\cdot(i+3)} \end{aligned}$$

and Lemma 2a we get

$$\begin{aligned} &\mathbf{E}\left\{ \frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)} \cdot \frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)} \right\} \\ &\quad -\mathbf{E}\left\{ \frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)} \right\} \\ &\quad \cdot \mathbf{E}\left\{ \frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(z)}(X_j)} \right\} \\ &= \mathbf{E}\left\{ \frac{1}{\left(1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l,r\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)\right)^2} \right\} \\ &\quad - \left(\mathbf{E}\left\{ \frac{1}{1+\sum_{j\in\{1,\ldots,n_k\}\setminus\{l\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)} \right\} \right)^2 \\ &= \sum_{i=0}^{n_k-2} \frac{1}{(1+i)^2} \cdot \binom{n_k-2}{i} p_1^i \cdot (1-p_1)^{n_k-2-i} - \left(\frac{1-(1-p_1)^{n_k}}{n_k \cdot p_1} \right)^2 \\ &\leq \sum_{i=0}^{n_k-2} \frac{1}{(i+1)\cdot(i+2)} \cdot \binom{n_k-2}{i} p_1^i \cdot (1-p_1)^{n_k-2-i} \\ &\quad + \sum_{i=0}^{n_k-2} \frac{3}{(i+1)\cdot(i+2)\cdot(i+3)} \cdot \binom{n_k-2}{i} p_1^i \cdot (1-p_1)^{n_k-2-i} \\ &\quad - \left(\frac{1-(1-p_1)^{n_k}}{n_k \cdot p_1} \right)^2 \\ &= \sum_{i=0}^{n_k-2} \frac{1}{(n_k-1)\cdot n_k \cdot p_1^2} \cdot \binom{n_k}{i+2} p_1^{i+2} \cdot (1-p_1)^{n_k-(i+2)} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=0}^{n_k-2} \frac{3}{(n_k-1)\cdot n_k\cdot(n_k+1)\cdot p_1^3}\cdot\binom{n_k+1}{i+3} p_1^{i+3}\cdot(1-p_1)^{n_k+1-(i+3)} \\
&-\left(\frac{1-(1-p_1)^{n_k}}{n_k\cdot p_1}\right)^2 \\
&= \frac{1}{(n_k-1)\cdot n_k\cdot p_1^2}\cdot\Big(1-(1-p_1)^{n_k}-n_k\cdot p_1\cdot(1-p_1)^{n_k-1}\Big) \\
&+\frac{3}{(n_k-1)\cdot n_k\cdot(n_k+1)\cdot p_1^3}\cdot\Big(1-(1-p_1)^{n_k+1}-(n_k+1)\cdot p_1\cdot(1-p_1)^{n_k} \\
&-\frac{1}{2}(n_k+1)\cdot n_k\cdot p_1^2\cdot(1-p_1)^{n_k-1}\Big)-\frac{1-2\cdot(1-p_1)^{n_k}+(1-p_1)^{2\cdot n_k}}{n_k^2\cdot p_1^2} \\
&\le \frac{1}{(n_k-1)\cdot n_k^2\cdot p_1^2}\cdot\left(n_k-n_k\cdot(1-p_1)^{n_k}\right)+\frac{3}{(n_k-1)\cdot n_k\cdot(n_k+1)\cdot p_1^3} \\
&-\frac{(n_k-1)-2\cdot(n_k-1)\cdot(1-p_1)^{n_k}}{(n_k-1)\cdot n_k^2\cdot p_1^2} \\
&= \frac{1+(n_k-2)\cdot(1-p_1)^{n_k}}{(n_k-1)\cdot n_k^2\cdot p_1^2}+\frac{3}{(n_k-1)\cdot n_k\cdot(n_k+1)\cdot p_1^3} \\
&\le \frac{5}{(n_k-1)\cdot n_k^2\cdot p_1^3},
\end{aligned}
$$

where the last inequality follows from

$$
(n_k-2)\cdot(1-p_1)^{n_k} \le \frac{n_k-2}{n_k\cdot p_1}\cdot(n_k\cdot p_1)\cdot e^{-n_k\cdot p_1} \le \frac{1}{p_1}\cdot\max_{z>0} z\cdot e^{-z} \le \frac{1}{p_1}
$$

and $p_1 \le 1$.

Consequently, we get by an application of the inequality of Cauchy–Schwarz and by Lemma 4a

$$
\begin{aligned}
&D_{k,l,r}^{(2)} \\
&\le \int\int_{A_{\mathcal{P}_{n_{k+1}}}(z)} \left(\mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X_{n_{k+1}})\cdot Y^*_{n_{k+1}}\right\}\right)^2 \\
&\cdot\frac{5}{(n_k-1)\cdot n_k^2\cdot\mu(A_{\mathcal{P}_{n_{k+1}}}(x))^3}\cdot\mu(\mathrm{d}x)\mu(dz) \\
&\le \int\int_{A_{\mathcal{P}_{n_{k+1}}}(z)} \mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X_{n_{k+1}})\cdot(Y^*_{n_{k+1}})^2\right\}\cdot\mathbf{E}\left\{I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X)\right\} \\
&\cdot\frac{5}{(n_k-1)\cdot n_k^2\cdot\mu(A_{\mathcal{P}_{n_{k+1}}}(z))^3}\cdot\mu(\mathrm{d}x)\mu(dz)
\end{aligned}
$$

$$= \int \mathbf{E}\left\{ I_{A_{\mathcal{P}^*_{n_{k+1}}}(z)}(X_{n_{k+1}}) \cdot (Y^*_{n_{k+1}})^2 \right\} \cdot \frac{5}{(n_k - 1) \cdot n_k^2 \cdot \mu(A_{\mathcal{P}_{n_{k+1}}}(z))} \mu(dz)$$

$$\leq c_8 \cdot \left(\frac{\max\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}}{\min\{h_{n_k}, h_{n_k+1}, \ldots, h_{n_{k+1}}\}} \right)^d \cdot \frac{\mathbf{E}\{(Y^*_{n_{k+1}})^2\}}{n_k^2 \cdot (n_k - 1)}.$$

Summarizing the above results, we get

$$\sum_{k=1}^{\infty} \sum_{\substack{l,r \in \{1,\ldots,n_{k+1}\} \\ l \neq r}} D_{k,l,r} \leq c_{10} \cdot \sum_{k=1}^{\infty} \frac{(\mathbf{E}\{Y\})^2 + \mathbf{E}\{(Y^*_{n_{k+1}})^2\}}{n_k} < \infty,$$

which completes the proof. □

Lemma 6 *Assume that the assumptions of Theorem 2 hold and set*

$$W_{n,i}(x) = \frac{K\left(\frac{x - X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x - X_j}{h_n}\right)}.$$

Then $W_{n,i}$ satisfies the assumptions $(A2)$, $(A3)$, $(A4)$ and $(A5)$ from Theorem 1.

Proof *Proof of $(A2)$:* Let $\delta > 0$ be arbitrary and by (8) choose $n_0 \in \mathbb{N}$ such that $\delta/h_n > r_2$ for $n \geq n_0$. Then we have for $n \geq n_0$

$$\sum_{i=1}^n W_{n,i}(x) \cdot I_{\{\|X_i - x\| > \delta\}} \leq \sum_{i=1}^n \frac{c_2 \cdot I_{S_{r_2}}\left(\frac{x - X_i}{h_n}\right) \cdot I_{\left\{\frac{\|X_i - x\|}{h_n} > \delta/h_n\right\}}}{\sum_{j=1}^n K\left(\frac{x - X_i}{h_n}\right)} = 0$$

for each $x \in \mathbb{R}^d$, which implies $(A2)$.
Proof of $(A3)$: It suffices to show $(A3')$ (compare Remark 2). W.l.o.g. we assume $c_2 = 1$. Then (15) and Lemma 5 yield $(A3')$.
Proof of $(A4)$: We have to show

$$Z_n := \int I_{\left\{\sum_{i=1}^n I_{S_{r_1}}\left(\frac{x - X_i}{h_n}\right) = 0\right\}} \mu(\mathrm{d}x) \to 0 \quad a.s.,$$

which follows from

$$\mathbf{E}\{Z_n\} \to 0 \quad (n \to \infty) \tag{24}$$

and

$$Z_n - \mathbf{E}\{Z_n\} \to 0 \quad a.s. \tag{25}$$

By assumption (9) and by arguing as on pages 75, 76 in Györfi et al. (2002), it is easy to see that (24) holds, and hence, it suffices to show (25). For this we use Lemma 4.2 in Kohler et al. (2003)—an Efron–Stein-type lemma for higher central moments—here for the fourth central moment.

Let X_1' be a d-dimensional random vector such that $X_1', X_1, \ldots, X_n$ are independent and identically distributed and set

$$Z_n' := \int I_{\left\{I_{S_{r_1}}\left(\frac{x-X_1'}{h_n}\right)+\sum_{i=2}^n I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)=0\right\}} \mu(\mathrm{d}x).$$

Then Lemma 4.2 in Kohler et al. (2003) implies that there exists a $c_{11} > 0$ such that for all $n \in \mathbb{N}$

$$\mathbf{E}\left\{|Z_n - \mathbf{E}\{Z_n\}|^4\right\} \leq c_{11} \cdot n^2 \cdot \mathbf{E}\left\{\left|Z_n - Z_n'\right|^4\right\}. \tag{26}$$

In case $I_{S_{r_1}}\left(\frac{x-X_1'}{h_n}\right) = I_{S_{r_1}}\left(\frac{x-X_1}{h_n}\right) = 0$ or $I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right) = 1$ for some $i \in \{2, \ldots, n\}$, we have

$$I_{\left\{\sum_{i=1}^n I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)=0\right\}} = I_{\left\{I_{S_{r_1}}\left(\frac{x-X_1'}{h_n}\right)+\sum_{i=2}^n I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)=0\right\}},$$

which implies

$$|Z_n - Z_n'| \leq \int \left(I_{S_{r_1}}\left(\frac{x-X_1'}{h_n}\right) + I_{S_{r_1}}\left(\frac{x-X_1}{h_n}\right)\right) \cdot I_{\left\{I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)=0 \quad (i=2,\ldots,n)\right\}} \mu(\mathrm{d}x).$$

Consequently,

$$\begin{aligned}
&\mathbf{E}\left\{\left|Z_n - Z_n'\right|^4\right\} \\
&\leq 16 \cdot \mathbf{E}\left\{\left(\int I_{S_{r_1}}\left(\frac{x-X_1}{h_n}\right) \cdot \prod_{i=2}^n \left(1 - I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)\right) \mu(\mathrm{d}x)\right)^4\right\} \\
&\leq 16 \cdot \mathbf{E}\left\{\left(\int I_{S_{r_1}}\left(\frac{x-X_1}{h_n}\right) \cdot \exp\left(-\sum_{i=2}^n I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)\right) \mu(\mathrm{d}x)\right)^4\right\} \\
&\leq 16 \cdot \mathbf{E}\left\{\left(\int I_{S_{r_1}}\left(\frac{x-X_1}{h_n}\right) \cdot \frac{1}{1 + \sum_{i=2}^n I_{S_{r_1}}\left(\frac{x-X_i}{h_n}\right)} \mu(\mathrm{d}x)\right)^4\right\},
\end{aligned}$$

where the last inequality followed from $(1+z) \cdot e^{-z} \leq 1$ $(z \in \mathbb{R}_+)$. Application of Lemma 5 yields

$$\mathbf{E}\left\{\left|Z_n - Z_n'\right|^4\right\} \leq \frac{c_{12}}{n^4},$$

from which we get the assertion by an application of the lemma of Borel–Cantelli, Markov inequality and inequality (26).

Proof of $(A5)$*:* Condition $(A5)$ follows directly from the definition of $W_{n.i}$ since K is nonnegative. □

Proof of Theorem 2. By Lemmas 1 and 6 the assumptions of Theorem 1 are satisfied. Application of Theorem 1 yields the assertion. □

Remark 4 The use of the Spiegelman and Sacks (1980) modification of the classic Nadaraya–Watson regression estimate allows an extension of Theorem 2 from boxed kernels to regular kernels in the sense of Definition 23.1 in Györfi et al. (2002), thus comprehending the Gaussian and also non-smooth kernels. The proof is similar to that of Theorem 2. Especially, one shows Lemma 1 for $W_{n,i}$ modified by inserting the additive constant 1 into the denominator, majorizes in the numerator the regular kernel by an infinite sum of weighted indicator functions in context of a bounded overlap cover of $\mathbb{R}^d$ with balls and uses Lemma 4. It remains an open question whether strong universal consistency holds for the classical Nadaraya–Watson regression estimate with Gaussian kernel.

3.3 Proof of Theorem 3

Analogously to the proof of Lemma 5, one can show

Lemma 7 *Assume that the assumptions of Theorem 3 are satisfied. Then constants* $c_{13}, c_{14} > 0$ *exists such that for each* $i \in \{1, \dots, n\}$

$$\mathbf{E}\left\{\left(\int \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)}\,\mu(dx)\right)^2\right\} \leq \frac{c_{13}}{n^2}$$

and

$$\mathbf{E}\left\{\left(\int \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)}\,\mu(dx)\right)^4\right\} \leq \frac{c_{14}}{n^4}.$$

Lemma 8 *Assume that the assumptions of Theorem 3 hold and set*

$$W_{n,i}(x) = \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)}{\sum_{j=1}^n I_{A_{\mathcal{P}_n}(x)}(X_j)}.$$

Then $W_{n,i}$ *satisfies the assumption* $(A1)$ *from Theorem 1.*

Proof of Lemma 8. Let Y be such that $Y \geq 0$ $a.s.$ We have to show for some $c_{15} > 0$

$$\limsup_{n\to\infty} \int \sum_{i=1}^n \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)\cdot Y_i}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)}\,\mu(\mathrm{d}x) \leq c_{15}\cdot \mathbf{E}Y. \tag{27}$$

Set $Y_i^* = Y_i \cdot I_{\{Y_i \leq i\}}$. We show next that (27) follows from

$$\limsup_{n\to\infty} \int \sum_{i=1}^n \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)\cdot Y_i^*}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)}\,\mu(\mathrm{d}x) \leq c_{15}\cdot \mathbf{E}Y. \tag{28}$$

To do this, we observe that we have as in the proof of Lemma 1 that $Y_n^* \neq Y_n$ holds at most for finitely many n. For each $i \in \mathbb{N}$, one has

$$\int \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)}{1 + \sum_{j \in \{1,\ldots,n\} \setminus \{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)} \mu(\mathrm{d}x) \to 0 \quad a.s.,$$

since Lemma 7 implies

$$\sum_{n=i}^{\infty} \mathbf{E} \left\{ \left(\int \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)}{1 + \sum_{j \in \{1,\ldots,n\} \setminus \{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)} \mu(\mathrm{d}x) \right)^2 \right\} \le c_{13} \cdot \sum_{n=i}^{\infty} \frac{1}{n^2} < \infty.$$

As in the proof of Lemma 1, we can conclude that (27) is implied by (28). In the sequel we show (28). Set $n_k = 2^k$ $(k \in \mathbb{N})$. Since the partitions $\mathcal{P}_n$ are nested, we have for any $r \le s$ and all $x \in \mathbb{R}^d$

$$A_{\mathcal{P}_r}(x) \supseteq A_{\mathcal{P}_s}(x),$$

from which we get for any $n \in \{n_k, n_k + 1, \ldots, n_{k+1}\}$

$$\begin{aligned} &\sum_{i=1}^{n} \frac{I_{A_{\mathcal{P}_n}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n\} \setminus \{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)} \\ &\quad \le \sum_{i=1}^{n_{k+1}} \frac{I_{A_{\mathcal{P}_n}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_n}(x)}(X_j)} \\ &\quad \le \sum_{i=1}^{n_{k+1}} \frac{I_{A_{\mathcal{P}_{n_k}}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)}. \end{aligned}$$

Hence, in order to show (28) it suffices to show

$$\limsup_{k \to \infty} \int \sum_{i=1}^{n_{k+1}} \frac{I_{A_{\mathcal{P}_{n_k}}(x)}(X_i) \cdot Y_i^*}{1 + \sum_{j \in \{1,\ldots,n_k\} \setminus \{i\}} I_{A_{\mathcal{P}_{n_{k+1}}}(x)}(X_j)} \mu(\mathrm{d}x) \le c_{15} \cdot \mathbf{E}Y.$$

By the assumptions of Theorem 3 each set in the partition $\mathcal{P}_{n_k}$ is contained in the union of at most L sets of the partition $\mathcal{P}_{n_{k+1}}$. Consequently, we get the assertion above as in the proof of Lemma 1. □

Lemma 9 *Assume that the assumptions of Theorem 3 hold and set*

$$W_{n,i}(x) = \frac{I_{A_{\mathcal{P}_n}(x)}(X_i)}{\sum_{j=1}^{n} I_{A_{\mathcal{P}_n}(x)}(X_j)}.$$

Then $W_{n,i}$ satisfies the assumptions (A2), (A3), (A4) and (A5) from Theorem 1.

Proof Proof of $(A2)$*:* Let $\epsilon > 0$ and $x \in \mathbb{R}^d$ be arbitrary. By assumption (11) there exists $n_0 \in \mathbb{N}$ such that $\mathrm{diam}(A_{\mathcal{P}_n}(x)) < \epsilon$ for $n \geq n_0$. Then we have for any $n \geq n_0$

$$\sum_{i=1}^{n} W_{n,i}(x) \cdot I_{\{\|X_i - x\| > \epsilon\}} = \sum_{i=1}^{n} \frac{I_{A_{\mathcal{P}_n}(x)}(X_i) \cdot I_{\{\|X_i - x\| > \epsilon\}}}{\sum_{j=1}^{n} I_{A_{\mathcal{P}_n}(x)}(X_j)} = 0,$$

which implies $(A2)$.
Proof of $(A3)$*:* Lemma 7 yields $(A3')$, which implies $(A3)$ (compare Remark 2).
Proof of $(A4)$*:* Analogously to the proof of $(A4)$ in Lemma 6.
Proof of $(A5)$*:* Condition $(A5)$ trivially follows from the definition of the weights. □

Proof of Theorem 3. By Lemmas 8 and 9 the assumptions of Theorem 1 are satisfied. Application of Theorem 1 yields the assertion. □

Acknowledgements The authors thank Editors and Reviewers for their comments, which led to a generalization of a result and to an improvement of the presentation.

References

Biau, G., Devroye, L. (2015). *Lectures on the nearest neighbor method*. Heidelberg: Springer.

Devroye, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates. *Annals of Statistics*, *9*, 1310–1319.

Devroye, L., Györfi, L., Krzyżak, A., Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics*, *22*, 1371–1385.

Devroye, L., Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference*, *23*, 71–82.

Devroye, L. P., Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics*, *8*, 231–239.

Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, *14*, 153–158.

Etemadi, N. (1981). An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *55*, 119–122.

Greblicki, W., Krzyżak, A., Pawlak, M. (1984). Distribution-free pointwise consistency of kernel regression estimate. *Annals of Statistics*, *12*, 1570–1575.

Györfi, L. (1991). Universal consistencies of a regression estimate for unbounded regression functions. In G. Roussas (Ed.), *Nonparametric functional estimation and related topics* (pp. 329–338)., NATO ASI series Dordrecht: Kluwer.

Györfi, L., Walk, H. (1996). On the strong universal consistency of a series type regression estimate. *Mathematical Methods of Statistics*, *5*, 332–342.

Györfi, L., Walk, H. (1997). On the strong universal consistency of a recursive regression estimate by Pál Révész. *Statistics and Probability Letters*, *31*, 177–183.

Györfi, L., Kohler, M., Krzyżak, A., Walk, H. (2002). *A distribution-free theory of nonparametric regression*. New York: Springer.

Györfi, L., Kohler, M., Walk, H. (1998). Weak and strong universal consistency of semi-recursive partitioning and kernel regression estimates. *Statistics & Decisions*, *16*, 1–18.

Irle, A. (1997). On consistency in nonparametric estimation under mixing conditions. *Journal of Multivariate Analysis*, *60*, 123–147.

Kohler, M. (1997). On the universal consistency of a least squares spline regression estimator. *Mathematical Methods of Statistics*, *6*, 349–364.

Kohler, M. (1999). Universally consistent regression function estimation using hierarchical B-splines. *Journal of Multivariate Analysis*, *67*, 138–164.

Kohler, M. (2003). Universal consistency of local polynomial kernel regression estimates. *Annals of the Institute of Statistical Mathematics*, *54*, 879–899.

Kohler, M., Krzyżak, A. (2001). Nonparametric regression estimation using penalized least squares. *IEEE Transactions on Information Theory*, *47*, 3054–3058.
Kohler, M., Krzyżak, A., Walk, H. (2003). Strong consistency of automatic kernel regression estimates. *Annals of the Institute of Statistical Mathematics*, *55*, 287–308.
Kozek, A. S., Leslie, J. R., Schuster, E. F. (1998). On a universal strong law of large numbers for conditional expectations. *Bernoulli*, *4*, 143–165.
Lugosi, G., Zeger, K. (1995). Nonparametric estimation via empirical risk minimization. *IEEE Transactions on Information Theory*, *41*, 677–687.
McDiarmid, C. (1989). On the method of bounded differences. In J. Siemons (Ed.), *Surveys in combinatorics* (pp. 148–188). Cambridge: Cambridge University Press.
Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, *9*, 141–142.
Spiegelman, C., Sacks, J. (1980). Consistent window estimation in nonparametric regression. *Annals of Statistics*, *8*, 240–246.
Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics*, *5*, 595–645.
Tukey, J. W. (1947). Nonparametric estimation II. Statistically equivalent blocks and tolerance regions—The continuous case. *Annals of Mathematical Statistics*, *18*, 529–539.
Walk, H. (2001). Strong universal pointwise consistency of recursive regression estimates. *Annals of the Institute of Statistical Mathematics*, *53*, 691–707.
Walk, H. (2002). Almost sure convergence properties of Nadaraya–Watson regression estimates. In M. Dror, P. L'Ecuyer, F. Szidarovszky (Eds.), *Modeling uncertainty: An examination of stochastic theory, methods, and applications (S. Yakowitz memorial volume)* (pp. 201–223). Dordrecht: Kluwer.
Walk, H. (2005). Strong universal consistency of smooth kernel regression estimation. *Annals of the Institute of Statistical Mathematics*, *57*, 665–685.
Walk, H. (2008). A universal strong law of large numbers for conditional expectations via nearest neighbors. *Journal of Multivariate Analysis*, *99*, 1035–1050.
Watson, G. S. (1964). Smooth regression analysis. *Sankhya Series A*, *26*, 359–372.