Ann. Inst. Statist. Math.
Vol. 57, No. 4, 665–685 (2005)

STRONG UNIVERSAL CONSISTENCY OF SMOOTH KERNEL REGRESSION ESTIMATES

HARRO WALK

Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany,
e-mail: walk@mathematik.uni-stuttgart.de

(Received April 16, 2003; revised October 20, 2004)

Abstract. The paper deals with kernel estimates of Nadaraya-Watson type for a regression function with square integrable response variable. For usual bandwidth sequences and smooth nonnegative kernels, e.g., Gaussian and quartic kernels, strong L_2-consistency is shown without any further condition on the underlying distribution. The proof uses a Tauberian theorem for Cesàro summability.

Key words and phrases: Nonparametric regression estimation, kernel estimate of Nadaraya and Watson, square integrability, strong and weak universal consistency, Efron-Stein inequality, covering, Tauberian theorem.

1. Introduction

Let X be a d-dimensional random vector with distribution μ and let Y be a real random variable with $EY^2 < \infty$. The regression function $m : \mathbb{R}^d \to \mathbb{R}$ defined by $m(x) := E(Y \mid X = x)$ is to be estimated on the basis of a sequence $(X_1, Y_1), (X_2, Y_2), \dots$ of independent identically distributed (i.i.d.) copies of (X, Y). No further assumptions on the distribution of (X, Y) will be required. For given values $(x_1, y_1), \dots, (x_n, y_n)$ of $(X_1, Y_1), \dots, (X_n, Y_n)$ an estimate of $m(x)$ will be denoted by $m_n(x, x_1, y_1, \dots, x_n, y_n) =: m_n(x), x \in \mathbb{R}^d$.

A sequence of estimates (m_n) is called *strongly* [*weakly*] *universally consistent* if

$$\int |m_n(x) - m(x)|^2 \mu(dx) \to 0 \quad \text{almost surely (a.s.)} \tag{1.1}$$

$$\left[E \int |m_n(x) - m(x)|^2 \mu(dx) \to 0\right] \tag{1.2}$$

for all (X, Y) with $EY^2 < \infty$.

Stone (1977) showed that there exist weakly universally consistent estimates, namely the nearest neighbor estimates. For suitably defined nearest neighbor estimates Devroye *et al.* (1994) showed strong universal consistency. For recursive series estimates, for recursive kernel estimates (Révész (1973)) and for semi-recursive kernel estimates (Devroye and Wagner (1980*b*)) and semi-recursive partitioning estimates, weak and strong universal consistency was proved by Györfi and Walk (1996, 1997) and Györfi *et al.* (1998). For suitably defined least squares estimates in function spaces Lugosi and Zeger (1995) and Kohler (1997, 1999) showed strong universal consistency using especially series estimates and spline estimates, respectively. As a reference for further consistency results in nonparametric regression estimation we mention Györfi *et al.* (2002).

The kernel estimate

$$\widehat{m}_n(x) := \frac{\sum_{i=1}^n Y_i K_{h_n}(x - X_i)}{\sum_{i=1}^n K_{h_n}(x - X_i)}, \quad \text{where} \quad K_{h_n} := K\left(\frac{\cdot}{h_n}\right) \tag{1.3}$$

($\frac{0}{0} := 0$) with bounded Lebesgue-integrable kernel $K : \mathbb{R}^d \to \mathbb{R}_+$ and positive bandwidths $h_n \to 0$, was introduced by Nadaraya (1964) and Watson (1964). Devroye and Wagner (1980*a*) and Spiegelman and Sacks (1980) showed weak universal consistency for K with compact support and $K \geq bI_{S_{0,r}}$ with some $b > 0$, $r > 0$, if $nh_n^d \to \infty$. (I denotes an indicator function and $S_{x,r}$ denotes the closed sphere around x in $\mathbb{R}^d$ with radius $r > 0$.) Strong consistency μ-almost everywhere for suitable kernels and bandwidth sequences was treated by Devroye (1981), Greblicki *et al.* (1984), Stute (1986), Kozek *et al.* (1998), Walk (2001) (for (semi-)recursive estimates), and other authors. In this paper we consider the modification

$$m_n(x) := \frac{\sum_{i=1}^n Y_i K_{h_n}(x - X_i)}{\max\{\delta, \sum_{i=1}^n K_{h_n}(x - X_i)\}}, \qquad x \in \mathbb{R}^d, \tag{1.4}$$

of the Nadaraya-Watson estimate $\widehat{m}_n(x)$, for fixed $\delta > 0$ (see Spiegelman and Sacks (1980)). For $1 \geq \delta > 0$, m_n coincides with $\widehat{m}_n$ in the case of the naive kernel $K = I_{S_{0,r}}$. The estimation function $m_n(\cdot)$ is continuous for continuous kernel. For K sufficiently smooth (e.g., Gaussian kernel $K(x) = e^{-\|x\|^2}$ or quartic kernel $K(x) = (1 - \|x\|^2)^2 I_{S_{0,1}}(x)$, $\|\cdot\|$ denoting the Euclidean norm) and usual bandwidth sequence (mainly $h_n = n^{-\gamma}$, $0 < \gamma d < 1$), we shall show strong universal consistency of (m_n) (Theorem 2.1). Whether strong universal consistency also holds for the Epanechnikov kernel K_E with $K_E(x) = (1 - \|x\|^2) I_{S_{0,1}}(x)$ or the naive kernel $K = I_{S_{0,r}}$, remains an open problem. The choice of bandwidths, especially its order, is important for the rate of convergence under regularity conditions (see Stone (1980, 1982), Krzyżak and Pawlak (1987), Härdle (1990, Chapter 4), Pawlak (1991), Györfi *et al.* (2002) and the literature cited there). For $d = 1$ with a two times differentiable m and existence of a differentiable density of X, in view of a small asymptotic mean square error (MSE) the ratio

$$\frac{(\int K(u)du)^{-2} (\int K(u)^2 du)^{4/5} (\int u^2 K(u) du)^{2/5}}{(\int K_E(u)du)^{-2} (\int K_E(u)^2 du)^{4/5} (\int u^2 K_E(u) du)^{2/5}}$$

with here optimal Epanechnikov kernel K_E is of interest (see e.g. Härdle (1990), pp. 29, 77, 100, 134, 138). For the Gaussian kernel the ratio is 1.041, while for the quartic kernel the ratio is 1.005.

Devroye and Krzyżak (1989) showed strong consistency of $(\widehat{m}_n)$ defined by (1.3), if Y is bounded, for rather general, so-called regular, kernels and bandwidth sequences (h_n) satisfying only $0 < h_n \to 0$, $nh_n^d \to \infty$. From this, it is easy to derive the corresponding result for (m_n) defined by (1.4). A lemma of Györfi (1991) formulated for general estimates of local averaging type (see Lemma 3.1) guarantees strong universal consistency of (m_n), if

$$\limsup \int |m_n(x)| \mu(dx) \leq cE|Y| \qquad \text{a.s.} \tag{1.5}$$

can be established with some constant $c > 0$ for all (X, Y) with $E|Y| < \infty$ (instead of $EY^2 < \infty$). Thus the crucial step in the proof of Theorem 2.1 is to show (1.5). This will be done in the slightly sharpened form

$$\int \sum_{i=1}^{n} \frac{Y_i K_{h_n}(x - X_i)}{1 + \sum_{j \in \{1,\dots,n\} \setminus \{i\}} K_{h_n}(x - X_j)} \mu(dx) \to EY \qquad \text{a.s.},$$

if the regular kernel K is choosen sufficiently smooth and (h_n) is of usual form (Lemma 3.13). In the proof of this lemma we first show a.s. convergence of the corresponding sequence of arithmetic means (Cesàro summability) and then use a Tauberian argument from summability theory (Lemmas 3.2 and 3.3) to obtain a.s. convergence. An important tool for both steps is the Efron-Stein (1981) inequality in Steele's (1986) version, but specialized for a symmetric function of n i.i.d. random vectors (see Lemma 3.4) together with a covering lemma of Devroye and Krzyżak (1989) (Lemma 3.6). The smoothness condition on K is needed for the application of the Tauberian Lemma 3.3. It should be mentioned that in the (trivial) case that μ is concentrated on a single point, the argument leads to a simple proof (without use of the Efron-Stein inequality) of Kolmogorov's strong law of large numbers for i.i.d. integrable real random variables (see Walk (2005)).

In Section 2 the result (Theorem 2.1) is formulated. Section 3 contains its proof together with several lemmas and their proofs.

2. Result

First smooth and regular kernels will be defined.

DEFINITION 2.1. Let H be a continuously differentiable nonincreasing function on $\mathbb{R}_+$ with $0 < H(0) \leq 1$ and $\int H(s)s^{d-1}ds < \infty$ such that R with $R(s) := s^2 H'(s)^{(2)}/H(s)$, $s \geq 0$ $(0/0 := 0)$, is bounded, piecewise continuous and, for s sufficiently large, nonincreasing, with $\int R(s)s^{d-1}ds < \infty$. The kernel $K : \mathbb{R}^d \to \mathbb{R}_+$ with $K(x) := H(\|x\|)$ shall be called a *smooth kernel.*

In this sense the Gaussian kernel $K = H(\|\cdot\|)$ with $H(s) = e^{-s^2}$ and the quartic kernel $K = H(\|\cdot\|)$ with $H(s) = (1 - s^2)^2 I_{[0,1]}(s)$ are smooth kernels. The smooth kernels are special regular kernels defined now.

DEFINITION 2.2. (Devroye and Krzyżak (1989)) The measurable kernel $K : \mathbb{R}^d \to \mathbb{R}_+$ is called *regular* if there are a sphere $S_{0,r}$ and a constant $b > 0$ such that $1 \geq K(x) \geq bI_{S_{0,r}}(x)$, $x \in \mathbb{R}^d$, and $\int \sup_{u \in S_{x,r}} K(u)dx < \infty$.

Let (h_n) be a bandwidth sequence satisfying $h_n \downarrow 0$, $nh_n^d \to \infty$, $h_n - h_{n+1} = O(h_n/n)$, e.g. $h_n = n^{-\gamma}$ with $0 < \gamma d < 1$. For a measurable kernel $K : \mathbb{R}^d \to \mathbb{R}_+$ set $K_{h_n} := K(\cdot/h_n)$. Fix $\delta \in (0, 1]$. Now define m_n by (1.4).

The following theorem states strong universal consistency of (m_n) for smooth kernels.

THEOREM 2.1. *Let K be a smooth kernel, (h_n) as before and (m_n) defined by (1.4). Assume $EY^2 < \infty$. Then (1.1) holds.*

In a similar but easier way (without a Tauberian argument) weak universal consistency, i.e., (1.2) can be shown for (m_n) defined by (1.4) with regular kernel K and $0 < h_n \to 0$, $nh_n^d \to \infty$, while Spiegelman and Sacks (1980) and Devroye and Wagner (1980a) showed the corresponding result for (m_n) and $(\hat{m}_n)$, respectively, with bounded support of K.

3. Proofs

In view of the proof of Theorem 2.1 we formulate several (essentially) known theorems as Lemmas 3.1–3.7. To make the paper more self-contained we shall prove some of them. As further tools Lemmas 3.8–3.13 will be proved. Throughout this section we assume $Y \geq 0$ and $Y_n \geq 0$, further $0 < \delta \leq 1$, without loss of generality. It is pointed out in the lemmas, which integrability or boundedness assumptions on Y and Y_n are used respectively.

The first lemma is useful for local averaging regression estimates to extend consistencies for bounded Y's to unbounded Y's. It is due to Györfi (1991) and can be proved by truncation of Y (see also Györfi *et al.* (2002), pp. 466, 467).

LEMMA 3.1. *Let m_n be of the (more general) form*

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

with measurable nonnegative weights $W_{n,i}(x) = W_{n,i}(x, X_1, \dots, X_n)$ satisfying

$$\sum_{i=1}^{n} W_{n,i}(x) \leq 1. \tag{3.1}$$

a) *Assume* (1.1) *for all (X, Y) with bounded Y. Further assume that there is a constant c such that* (1.5) *holds for all (X, Y) with $EY < \infty$. Then (m_n) is strongly universally consistent.*

b) *Assume $\int m_n(x)\mu(dx) \to EY$ a.s. for all (X, Y) with bounded Y. Further assume that there is a constant c such that* (1.5) *holds for all (X, Y) with $EY < \infty$. Then $\int m_n(x)\mu(dx) \to EY$ a.s. for all (X, Y) with $EY < \infty$.*

The next lemma is a classical Tauberian theorem of R. Schmidt in summability theory (see Hardy (1949), Theorem 68, and Zeller and Beekmann (1970), 52 II, with references). For a sequence of real numbers with convergence of the corresponding sequence of arithmetic means (Cesàro summability), convergence is established under an additional condition (Tauberian condition).

LEMMA 3.2. *If the sequence $(s_n)_{n \in \mathbb{N}}$ of real numbers satisfies*

$$t_n := (s_1 + \cdots + s_n)/n \to 0 \tag{3.2}$$

and

$$\liminf (s_N - s_M) \geq 0 \quad \textit{for} \quad N > M \to \infty,\ N/M \to 1, \tag{3.3}$$

then

$$s_n \to 0. \tag{3.4}$$

PROOF. With $N = N(M) > M$ such that $N/M \to 1$ $(M \to \infty)$ sufficiently slow and with

$$s_M = \frac{1}{N-M} \sum_{k=M+1}^{N} s_k - \frac{1}{N-M} \sum_{k=M+1}^{N} (s_k - s_M) =: A_{M,N} - B_{M,N}$$

one has

$$\begin{aligned} A_{M,N} &= \frac{1}{N-M} \left[\sum_{k=1}^{N} s_k - \sum_{k=1}^{M} s_k \right] \\ &= \frac{N/M}{N/M - 1} t_N - \frac{1}{N/M - 1} t_M \to 0 \qquad \text{(by (3.2))}, \\ B_{M,N} &\geq \min_{k=M+1,\dots,N} (s_k - s_M) = s_{k(M,N)} - s_M \end{aligned}$$

with $k(M,N) \in \{M+1, \dots, N\}$, thus by (3.3) with $k(M,N)$ instead of N

$$\limsup s_M = -\liminf B_{M,N} \leq 0. \tag{3.5}$$

Correspondingly, with $L = L(M) < M$ such that $L/M \to 1$ $(M \to \infty)$ sufficiently slow and with

$$s_M = -\frac{1}{M-L} \sum_{k=L+1}^{M} s_k + \frac{1}{M-L} \sum_{k=L+1}^{M} (s_M - s_k) = -A_{L,M} + B_{L,M}$$

one has

$$\begin{aligned} &A_{L,M} \to 0 \qquad \text{(by (3.2))}, \\ &\liminf s_M \geq \liminf B_{L,M} \geq 0 \qquad \text{(by (3.3))}. \end{aligned} \tag{3.6}$$

(3.5) and (3.6) yield (3.4). □

We reformulate Lemma 3.2 in a stochastic version as Lemma 3.3 which will be directly applied in the proof of Lemma 3.13.

LEMMA 3.3. *Let U_n, J_n, W_n, Z_n, V be real random variables and let c_n be real numbers such that*

$$U_{n+1} - U_n = J_n + c_n + W_n + Z_n, \qquad n \in \mathbb{N}.$$

Assume $\frac{1}{N} \sum_{n=1}^{N} U_n \to V$ a.s., further $J_n \geq 0$ $(n \in \mathbb{N})$,

$$\lim \sum_{n=M}^{N} c_n = 0, \qquad \textit{if} \quad N > M \to \infty, \; N/M \to 1,$$

$$\sum n W_n^2 < \infty \qquad \textit{a.s.}, \qquad \sum Z_n \qquad \textit{a.s. convergent.}$$

Then $U_n \to V$ a.s.

PROOF. Let $(\Omega, \mathcal{A}, \mathcal{P})$ be the underlying probability space, and set

$$\Omega' := \left\{ \omega \in \Omega; \frac{1}{N} \sum_{n=1}^{N} U_n(\omega) \to V(\omega), \sum n W_n^2(\omega) < \infty, \sum Z_n(\omega) \text{ convergent} \right\}.$$

Then $P(\Omega') = 1$. Choose $\omega \in \Omega'$. For arbitrary $N > M \to \infty$, $N/M \to 1$, we have

$$\begin{aligned} \left| \sum_{n=M}^{N} W_n(\omega) \right| &\le \sum_{n=M}^{N} \frac{1}{\sqrt{n}} \sqrt{n} |W_n(\omega)| \\ &\le \left(\sum_{n=M}^{N} \frac{1}{n} \right)^{1/2} \left(\sum_{n=M}^{N} n W_n(\omega)^2 \right)^{1/2} \to 0, \end{aligned}$$

$$\sum_{n=M}^{N} Z_n(\omega) \to 0,$$

thus

$$\liminf (U_{N+1}(\omega) - U_M(\omega)) = \liminf \sum_{n=M}^{N} (J_n(\omega) + c_n + W_n(\omega) + Z_n(\omega)) \ge 0.$$

This together with

$$\frac{1}{N} \sum_{n=1}^{N} U_n(\omega) \to V(\omega) \qquad (N \to \infty)$$

yields $U_n(\omega) \to V(\omega)$ $(n \to \infty)$ by Lemma 3.2. □

In the following we specialize the Efron-Stein (1981) inequality in Steele's (1986) version (see Györfi *et al.* (2002) for further references) to the case of a symmetric statistic of i.i.d. random vectors.

LEMMA 3.4. *Let $Z_1, \ldots, Z_n, \tilde{Z}_n$ be i.i.d. random variables with values in some Borel set $A \subset B^m$, and let the function $f : A^n \to \mathbb{R}$ be measurable and symmetric (i.e., the function values are not changed by a permutation of the arguments). If $f(Z_1, \ldots, Z_n)$ is square integrable, then*

$$\operatorname{Var} f(Z_1, \ldots, Z_n) \le \frac{1}{2} n E |f(Z_1, \ldots, Z_n) - f(Z_1, \ldots, Z_{n-1}, \tilde{Z}_n)|^2.$$

Further we mention a lemma which is well-known from the classical Kolmogorov proof of the strong law of large numbers for i.i.d. integrable real random variables (see, e.g., Loève (1977), section 17, as a reference). Here and later on, for $Y_n \ge 0$ let $Y_n^{[L]} := Y_n I_{[Y_n \le L]}$ denote the truncation at $L > 0$.

LEMMA 3.5. *For identically distributed random variables Y_n (≥ 0) with $EY_n < \infty$ one has*

a) *a.s. $Y_n = Y_n^{[n]}$ from some random index on,*

b) $\sum_{n=1}^{\infty} \frac{1}{n^2} E(Y_n^{[n]})^2 < \infty$.

The following covering lemma is due to Devroye and Krzyżak (1989), there with kernel $T = K$. Recall the notation $K_h := K(\cdot/h)$, correspondingly $T_h := T(\cdot/h)$ for $h > 0$.

LEMMA 3.6. *If the kernels T and K are regular, then there exists a finite constant $\rho = \rho(T, K)$ depending only upon T and K such that, for any $u \in \mathbb{R}^d$, $h > 0$, and probability measure μ,*

$$\int \frac{T_h(x-u)}{\int K_h(x-z)\mu(dz)} \mu(dx) \leq \rho.$$

PROOF. The proof follows the argument of Devroye and Krzyżak (1989). A natural number k_1 depending only on d, and a sequence $(x_i)_{i \in \mathbb{N}}$ in $\mathbb{R}^d$ exist such that the family $\{S_{x_i,r/2}; i \in \mathbb{N}\}$ of balls covers $\mathbb{R}^d$ where each $x \in \mathbb{R}^d$ gets covered at most k_1 times. One notices that $x \in S_{x_i,r/2}$ implies $S_{x_i,r/2} \subset S_{x,r}$, and obtains for the regular kernel T

$$\begin{aligned}
&\sum_{i=1}^{\infty} \sup_{z \in S_{x_i,r/2}} T(z) \\
&\quad = \sum_{i=1}^{\infty} \frac{1}{\operatorname{vol}(S_{0,r/2})} \int_{S_{x_i,r/2}} \sup_{z \in S_{x_i,r/2}} T(z)dx \\
&\quad \leq \frac{1}{\operatorname{vol}(S_{0,r/2})} \int \sum_{i=1}^{\infty} I_{S_{x_i,r/2}}(x) \sup_{z \in S_{x,r}} T(z)dx \\
&\quad \leq \frac{k_1}{\operatorname{vol}(S_{0,r/2})} \int \sup_{z \in S_{x,r}} T(z)dx \\
&\quad \leq k_2 < \infty
\end{aligned}$$

with k_2 depending only on d and T. For arbitrary $u \in \mathbb{R}$ one has

$$\begin{aligned}
T_h(x-u) &\leq \sum_{i=1}^{\infty} I_{S_{u+hx_i,rh/2}}(x) \sup_{t \in S_{u+hx_i,rh/2}} T_h(t-u) \\
&= \sum_{i=1}^{\infty} I_{S_{u+hx_i,rh/2}}(x) \sup_{z \in S_{x_i,r/2}} T(z)
\end{aligned}$$

and notices that $x \in S_{u+hx_i,rh/2}$ implies

$$\int K_h(x-z)\mu(dz) \geq b\mu(S_{x,rh}) \geq b\mu(S_{u+hx_i,rh/2})$$

for the regular kernel K. Therefore

$$\begin{aligned}
&\int \frac{T_h(x-u)}{\int K_h(x-z)\mu(dz)} \mu(dx) \\
&\quad \leq \sum_{i=1}^{\infty} \int_{S_{u+hx_i,rh/2}} \frac{1}{b\mu(S_{u+hx_i,rh/2})} \mu(dx) \sup_{z \in S_{x_i,r/2}} T(z)
\end{aligned}$$

$$\leq \frac{1}{b}\sum_{i=1}^{\infty} \sup_{z\in S_{x_i,r/2}} T(z)$$
$$\leq \frac{k_2}{b}$$

for all $u \in \mathbb{R}^d$. $\square$

Also the following result is due to Devroye and Krzyżak (1989).

LEMMA 3.7. *Let K be a regular kernel and let (h_n) satisfy $0 < h_n \to 0$, $nh_n^d \to \infty$. If Y is bounded, then*

$$\int |\hat{m}_n(x) - m(x)|\mu(dx) \to 0 \qquad a.s.$$

LEMMA 3.8. *Let $K : \mathbb{R}^d \to \mathbb{R}$, $T : \mathbb{R}^d \to \mathbb{R}$ be regular kernels. There is a $c \in \mathbb{R}_+$ such that for all $z \in \mathbb{R}^d$, $n \in \mathbb{N}$, $h > 0$ the following relations hold:*

a) $$E\left(\int \frac{T_h(x-z)}{1+\sum_{j=1}^n K_h(x-X_j)}\mu(dx)\right)^2 \leq \frac{c}{n^2},$$

b) $$E\left(\int \frac{T_h(x-z)K_h(x-X)}{[1+\sum_{j=1}^n K_h(x-X_j)]^2}\mu(dx)\right)^2 \leq \frac{c}{n^3},$$

c) $$E\int \frac{T_h(x-z)K_h(x-X)}{[1+\sum_{j=1}^n K_h(x-X_j)]^2}\mu(dx)\int \frac{K_h(\tilde{x}-X)}{1+\sum_{j=1}^n K_h(\tilde{x}-X_j)}\mu(d\tilde{x}) \leq \frac{c}{n^3},$$

d) $$E\left(\int \frac{T_h(x-z)EK_h(x-X)}{[1+\sum_{j=1}^n K_h(x-X_j)]^2}\mu(dx)\right)^2 \leq \frac{c}{n^4}.$$

PROOF. Let $n \geq 3$ w.l.o.g. We notice $0 \leq K \leq 1$, and use i.i.d. random vectors $X', \tilde{X}, X, X_1, X_2, \ldots$ and their exchangeability.

a) We obtain

$$E\left(\int \frac{T_h(x-z)}{1+\sum_{j=1}^n K_h(x-X_j)}\mu(dx)\right)^2$$
$$= E\frac{T_h(X'-z)T_h(\tilde{X}-z)}{[1+\sum_{j=1}^n K_h(X'-X_j)][1+\sum_{j=1}^n K_h(\tilde{X}-X_j)]}$$
$$\leq 4E\frac{T_h(X'-z)T_h(\tilde{X}-z)}{[2+\sum_{j=3}^n K_h(X'-X_j)][2+\sum_{j=3}^n K_h(\tilde{X}-X_j)]}$$
$$= 4E\frac{T_h(X_1-z)T_h(X_2-z)}{[2+\sum_{j=3}^n K_h(X_1-X_j)][2+\sum_{j=3}^n K_h(X_2-X_j)]} \qquad \text{(by exchangeability)}$$
$$\leq 4E\frac{T_h(X_1-z)T_h(X_2-z)}{\sum_{j=1}^n K_h(X_1-X_j)\sum_{j=1}^n K_h(X_2-X_j)}$$

$$= 4\frac{1}{n(n-1)}E\sum_{\substack{i,l\in\{1,\dots,n\}\\ i\neq l}}\frac{T_h(X_i-z)T_h(X_l-z)}{\sum_{j=1}^n K_h(X_i-X_j)\sum_{j=1}^n K_h(X_l-X_j)}$$

(by exchangeability)

$$\leq 8\rho^2\frac{1}{n^2}$$

the latter by Lemma 3.6 for the empirical measure based on $X_1,\dots,X_n$.

b) Similarly,

$$E\left(\int\frac{T_h(x-z)K_h(x-X)}{[1+\sum_{j=1}^n K_h(x-X_j)]^2}\mu(dx)\right)^2$$
$$= E\frac{T_h(X'-z)K_h(X'-X)}{[1+\sum_{j=1}^n K_h(X'-X_j)]^2}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X)}{[1+\sum_{j=1}^n K_h(\tilde{X}-X_j)]^2}$$
$$\leq 8E\frac{T_h(X'-z)}{2+\sum_{j=2}^n K_h(X'-X_j)}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X)}{[2+\sum_{j=2}^n K_h(\tilde{X}-X_j)]^2}$$
$$\leq 8E\frac{T_h(X'-z)}{1+\sum_{j=1}^n K_h(X'-X_j)}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X_1)}{[1+\sum_{j=1}^n K_h(\tilde{X}-X_j)]^2}$$
$$= \frac{8}{n}\sum_{i=1}^n E\frac{T_h(X'-z)}{1+\sum_{j=1}^n K_h(X'-X_j)}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X_i)}{[1+\sum_{j=1}^n K_h(\tilde{X}-X_j)]^2}$$

(by exchangeability)

$$\leq \frac{8}{n}E\left(\int\frac{T_h(x-z)}{1+\sum_{j=1}^n K_h(x-X_j)}\mu(dx)\right)^2$$
$$\leq \frac{c'}{n^3}$$

for some $c'\in\mathbb{R}_+$, the latter by a).

c) We obtain

$$E\int\frac{T_h(x-z)K_h(x-X)}{[1+\sum_{j=1}^n K_h(x-X_j)]^2}\mu(dx)\int\frac{K_h(\tilde{x}-X)}{1+\sum_{j=1}^n K_h(\tilde{x}-X_j)}\mu(d\tilde{x})$$
$$= E\frac{T_h(X'-z)K_h(X'-X)}{[1+\sum_{j=1}^n K_h(X'-X_j)]^2}\frac{K_h(\tilde{X}-X)}{1+\sum_{j=1}^n K_h(\tilde{X}-X_j)}$$
$$\leq 9E\frac{T_h(X'-z)K_h(X'-X)}{[3+\sum_{j=2}^n K_h(X'-X_j)]^2}\frac{K_h(\tilde{X}-X)}{1+\sum_{j=2}^n K_h(\tilde{X}-X_j)}$$
$$\leq 9E\frac{T_h(X'-z)K_h(X'-X)}{[2+\sum_{j=1}^n K_h(X'-X_j)]^2}\frac{K_h(X_1-X)}{\sum_{j=1}^n K_h(X_1-X_j)}$$
$$= 9\frac{1}{n}E\frac{T_h(X'-z)K_h(X'-X)}{[2+\sum_{j=1}^n K_h(X'-X_j)]^2}\sum_{i=1}^n\frac{K_h(X_i-X)}{\sum_{j=1}^n K_h(X_i-X_j)}$$

(by exchangeability)

$$\leq 9\rho\frac{1}{n}E\frac{T_h(X'-z)K_h(X'-X)}{[2+\sum_{j=2}^n K_h(X'-X_j)]^2}$$

(by Lemma 3.6 for the empirical measure based on $X_1,\ldots,X_n$)

$$\leq 9\rho\frac{1}{n}E\frac{T_h(X'-z)K_h(X'-X_1)}{[1+\sum_{j=1}^n K_h(X'-X_j)]^2}$$

$$= 9\rho\frac{1}{n^2}E\sum_{i=1}^n\frac{T_h(X'-z)K_h(X'-X_i)}{[1+\sum_{j=1}^n K_h(X'-X_j)]^2} \qquad \text{(by exchangeability)}$$

$$\leq 9\rho\frac{1}{n^2}E\frac{T_h(X_1-z)}{1+\sum_{j=2}^n K_h(X_1-X_j)}$$

$$\leq 9\rho\frac{1}{n^2}E\frac{T_h(X_1-z)}{\sum_{j=1}^n K_h(X_1-X_j)}$$

$$= 9\rho\frac{1}{n^3}E\sum_{i=1}^n\frac{T_h(X_i-z)}{\sum_{j=1}^n K_h(X_i-X_j)} \qquad \text{(by exchangeability)}$$

$$\leq 9\rho^2\frac{1}{n^3},$$

the latter as in a) by Lemma 3.6 for the empirical measure.

d)

$$E\left(\int\frac{T_h(x-z)EK_h(x-X)}{[1+\sum_{j=1}^n K_h(x-X_j)]^2}\mu(dx)\right)^2$$

$$= E\frac{T_h(X'-z)K_h(X'-X_{n+1})}{[1+\sum_{j=1}^n K_h(X'-X_j)]^2}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X_{n+2})}{[1+\sum_{j=1}^n K_h(\tilde{X}-X_j)]^2}$$

$$\leq 81E\frac{T_h(X'-z)K_h(X'-X_{n+1})}{[3+\sum_{j=1}^n K_h(X'-X_j)]^2}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X_{n+2})}{[3+\sum_{j=1}^n K_h(\tilde{X}-X_j)]^2}$$

$$\leq 81E\frac{T_h(X'-z)K_h(X'-X_{n+1})}{[1+\sum_{j=1}^{n+2} K_h(X'-X_j)]^2}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X_{n+2})}{[1+\sum_{j=1}^{n+2} K_h(\tilde{X}-X_j)]^2}$$

$$= \frac{81}{(n+2)(n+1)}$$

$$\cdot\sum_{\substack{i,l\in\{1,\ldots,n+2\}\\ i\neq l}} E\frac{T_h(X'-z)K_h(X'-X_i)}{[1+\sum_{j=1}^{n+2} K_h(X'-X_j)]^2}\frac{T_h(\tilde{X}-z)K_h(\tilde{X}-X_l)}{[1+\sum_{j=1}^{n+2} K_h(\tilde{X}-X_j)]^2}$$

(by exchangeability)

$$\leq \frac{81}{(n+2)(n+1)}E\left(\int\frac{T_h(x-z)}{1+\sum_{j=1}^{n+2} K_h(x-X_j)}\mu(dx)\right)^2$$

$$\leq \frac{c''}{n^4}$$

for some $c''\in\mathbb{R}_+$, the latter by a). □

LEMMA 3.9. *Let $K: R^d\to\mathbb{R}$ and $T:\mathbb{R}^d\to\mathbb{R}$ be regular kernels. For parts* b) *and* c) *below, assume that also T^2/K is a regular kernel. There is $c\in\mathbb{R}_+$ such that for* *all* $n \in \mathbb{N}$, $h > 0$ *the following relations hold*:

$$\text{a)}\quad E\left(\sum_{i=1}^{n-1}\int \frac{Y_i^{[n]}T_h(x-X_i)EK_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\right)^2 \le \frac{c}{n^2}E(Y_n^{[n]})^2,$$

$$\text{b)}\quad E\left(\sum_{i=1}^{n-1}\int \frac{Y_i^{[n]}T_h(x-X_i)K_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\right)^2 \le \frac{c}{n^2}E(Y_n^{[n]})^2,$$

$$\text{c)}\quad E\left(\sum_{i=1}^{n-1}\int \frac{Y_i^{[n]}T_h(x-X_i)K_h(x-X_n)EK_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^3}\mu(dx)\right)^2 \le \frac{c}{n^4}E(Y_n^{[n]})^2.$$

PROOF. Let $n \ge 4$ w.l.o.g.

a) By the Cauchy-Schwarz inequality the left-hand side is bounded by

$$\begin{aligned}
&(n-1)\sum_{i=1}^{n-1}E\left\{(Y_i^{[n]})^2\left(\int \frac{T_h(x-X_i)EK_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\right)^2\right\}\\
&\le (n-1)^2\int E((Y_1^{[n]})^2 \mid X_1=z)E\left(\int \frac{T_h(x-z)EK_h(x-X_n)}{[1+\sum_{j=2}^{n-1}K_h(x-X_j)]^2}\mu(dx)\right)^2\mu(dz)\\
&\le \frac{c}{n^2}\int E((Y_1^{[n]})^2 \mid X_1=z)\mu(dz) \qquad \text{(by Lemma 3.8 d))}\\
&= \frac{c}{n^2}E(Y_n^{[n]})^2
\end{aligned}$$

for some $c \in \mathbb{R}_+$.

b) The left-hand side is bounded by

$$\begin{aligned}
&\sum_{i=1}^{n-1}E\left(\int \frac{Y_i^{[n]}T_h(x-X_i)K_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\right)^2\\
&\qquad + \sum_{\substack{i,i'\in\{1,\dots,n-1\}\\ i\ne i'}} E\int \frac{Y_i^{[n]}T_h(x-X_i)K_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i,i'\}}K_h(x-X_j)]^2}\mu(dx)\\
&\qquad \cdot\int \frac{Y_{i'}^{[n]}T_h(\tilde{x}-X_{i'})K_h(\tilde{x}-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i,i'\}}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x})\\
&\quad = (n-1)H_n + (n-1)(n-2)I_n
\end{aligned}$$

with

$$\begin{aligned}
H_n &= E(Y_1^{[n]})^2\left(\int \frac{T_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=2}^{n-1}K_h(x-X_j)]^2}\mu(dx)\right)^2,\\
I_n &= EY_1^{[n]}Y_2^{[n]}\int \frac{T_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=3}^{n-1}K_h(x-X_j)]^2}\mu(dx)\int \frac{T_h(\tilde{x}-X_2)K_h(\tilde{x}-X_n)}{[1+\sum_{j=3}^{n-1}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x}).
\end{aligned}$$

We have

$$
\begin{aligned}
H_n &= \int E((Y_1^{[n]})^2 \mid X_1 = z) E\left(\left(\int \frac{T_h(x-z)K_h(x-X_n)}{[1+\sum_{j=2}^{n-1} K_h(x-X_j)]^2}\mu(dx)\right)^2\right)\mu(dz) \\
&\le \frac{c}{n^3}\int E((Y_1^{[n]})^2 \mid X_1 = z)\mu(dz) \qquad \text{(by Lemma 3.8 b))} \\
&= \frac{c}{n^3}E(Y_n^{[n]})^2
\end{aligned}
$$

for some $c \in \mathbb{R}_+$. Now we set $W = T^2/K, W_h = T_h^2/K_h$. By

$$
\begin{aligned}
ab &= \frac{a}{p}\frac{b}{q}pq \\
&\le \frac{1}{2}\frac{a^2}{p^2}pq + \frac{1}{2}\frac{b^2}{q^2}pq = \frac{1}{2}\frac{a^2}{p}q + \frac{1}{2}\frac{b^2}{q}p
\end{aligned}
$$

for $a, b, p, q \in \mathbb{R}_+$ with $p > 0$, $q > 0$ or $a = p = 0$ or $b = q = 0$, where $0/0 := 0$, we obtain

$$
\begin{aligned}
&Y_1^{[n]}Y_2^{[n]}T_h(x-X_1)T_h(\tilde{x}-X_2) \\
&\quad \le \frac{1}{2}(Y_1^{[n]})^2 W_h(x-X_1)K_h(\tilde{x}-X_2) + \frac{1}{2}(Y_2^{[n]})^2 W_h(\tilde{x}-X_2)K_h(x-X_1).
\end{aligned}
$$

Thus

$$
\begin{aligned}
I_n &\le \frac{1}{2}E(Y_1^{[n]})^2 \int \frac{W_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=3}^{n-1}K_h(x-X_j)]^2}\mu(dx) \int \frac{K_h(\tilde{x}-X_2)K_h(\tilde{x}-X_n)}{[1+\sum_{j=3}^{n-1}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x}) \\
&\quad + \frac{1}{2}E(Y_2^{[n]})^2 \int \frac{K_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=3}^{n-1}K_h(x-X_j)]^2}\mu(dx) \int \frac{W_h(\tilde{x}-X_2)K_h(\tilde{x}-X_n)}{[1+\sum_{j=3}^{n-1}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x}) \\
&= E(Y_1^{[n]})^2 \int \frac{W_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=3}^{n-1}K_h(x-X_j)]^2}\mu(dx) \int \frac{K_h(\tilde{x}-X_2)K_h(\tilde{x}-X_n)}{[1+\sum_{j=3}^{n-1}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x}) \\
&\hspace{25em} \text{(by exchangeability)} \\
&\le 16E(Y_1^{[n]})^2 \int \frac{W_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=2}^{n-1}K_h(x-X_j)]^2}\mu(dx) \int \frac{K_h(\tilde{x}-X_2)K_h(\tilde{x}-X_n)}{[1+\sum_{j=2}^{n-1}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x}) \\
&= \frac{16}{n-2}E(Y_1^{[n]})^2 \int \frac{W_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=2}^{n-1}K_h(x-X_j)]^2}\mu(dx) \\
&\quad \cdot \int \sum_{k=2}^{n-1} \frac{K_h(\tilde{x}-X_k)K_h(\tilde{x}-X_n)}{[1+\sum_{j=2}^{n-1}K_h(\tilde{x}-X_j)]^2}\mu(d\tilde{x}) \qquad \text{(by exchangeability)} \\
&\le \frac{16}{n-2}E(Y_1^{[n]})^2 \int \frac{W_h(x-X_1)K_h(x-X_n)}{[1+\sum_{j=2}^{n-1}K_h(x-X_j)]^2}\mu(dx) \int \frac{K_h(\tilde{x}-X_n)}{1+\sum_{j=2}^{n-1}K_h(\tilde{x}-X_j)}\mu(d\tilde{x}) \\
&\le \frac{c'}{n^4}\int E((Y_1^{[n]})^2 \mid X_1 = z)\mu(dz) \\
&\quad \text{(by conditioning with respect to } X_1 \text{ and by Lemma 3.8 c) with } T \text{ replaced by } W) \\
&= \frac{c'}{n^4}E(Y_n^{[n]})^2
\end{aligned}
$$

for some $c' \in \mathbb{R}_+$. This yields the assertion.

c) The left-hand side equals

$$
\begin{aligned}
&E \sum_{i,l\in\{1,\dots,n-1\}} Y_i^{[n]} \int \frac{T_h(x-X_i)K_h(x-X_n)K_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}} K_h(x-X_j)]^3}\mu(dx)\\
&\qquad \cdot Y_l^{(n)} \int \frac{T_h(\tilde{x}-X_l)K_h(\tilde{x}-X_n)K_h(\tilde{x}-X_{n+2})}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{l\}} K_h(\tilde{x}-X_j)]^3}\mu(d\tilde{x})\\
&\quad \le 64E \sum_{i,l\in\{1,\dots,n-1\}} Y_i^{[n]} \int \frac{T_h(x-X_i)K_h(x-X_n)K_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n-1,n+1\}\setminus\{i\}} K_h(x-X_j)]^3}\mu(dx)\\
&\qquad \cdot Y_l^{[n]} \int \frac{T_h(\tilde{x}-X_l)K_h(\tilde{x}-X_n)K_h(\tilde{x}-X_{n+2})}{[1+\sum_{j\in\{1,\dots,n-1,n+1\}\setminus\{l\}} K_h(\tilde{x}-X_j)]^3}\mu(d\tilde{x})\\
&\quad \le 64\frac{1}{n-2}E \sum_{i,l\in\{1,\dots,n-1\}}\\
&\qquad Y_i^{[n]} \int \frac{T_h(x-X_i)K_h(x-X_n)\sum_{j\in\{1,\dots,n-1,n+1\}\setminus\{i,l\}} K_h(x-X_j)}{[1+\sum_{j\in\{1,\dots,n-1,n+1\}\setminus\{i\}} K_h(x-X_j)]^3}\mu(dx)\\
&\qquad \cdot Y_l^{[n]} \int \frac{T_h(\tilde{x}-X_l)K_h(\tilde{x}-X_n)K_h(\tilde{x}-X_{n+2})}{[1+\sum_{j\in\{1,\dots,n-1,n+1\}\setminus\{l\}} K_h(\tilde{x}-X_j)]^3}\mu(d\tilde{x})\\
&\qquad\qquad \text{(by exchangeability, with } \{i,l\}=\{i\} \text{ for } i=l)\\
&\quad \le 64\cdot 32\frac{1}{(n-2)(n-1)}\\
&\qquad \cdot E\left(\sum_{i=1}^{n-1} Y_i^{[n]} \int \frac{T_h(x-X_i)K_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}} K_h(x-X_j)]^2}\mu(dx)\right)^2\\
&\qquad\qquad \text{(by simplifying the first fraction and by a repetition of the argument)}\\
&\quad \le \frac{c}{n^4}E(Y_n^{[n]})^2
\end{aligned}
$$

for some $c \in \mathbb{R}_+$, the latter by part b). □

LEMMA 3.10. *Let K, T and also T^2/K be regular kernels. There is a $c \in \mathbb{R}_+$ such that for all $n \in \mathbb{N}$, $h > 0$ the following relations hold:*

a) $$\operatorname{Var}\left(\int \sum_{i=1}^{n} \frac{Y_i^{[n]}T_h(x-X_i)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_h(x-X_j)}\mu(dx)\right) \le \frac{c}{n}E(Y_n^{[n]})^2,$$

b) $$\operatorname{Var}\left(\int \sum_{i=1}^{n} \frac{Y_i^{[n]}T_h(x-X_i)EK_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_h(x-X_j)]^2}\mu(dx)\right) \le \frac{c}{n^3}E(Y_n^{[n]})^2,$$

c) $$E\left(\sum_{i=1}^{n} \frac{Y_i^{[n]}T_h(x-X_i)[K_h(x-X_{n+1})-EK_h(x-X_{n+1})]}{[1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_h(x-X_j)]^2}\mu(dx)\right)^2 \le \frac{c}{n^2}E(Y_n^{[n]})^2.$$

PROOF. For the proof of parts a) and b) we use Lemma 3.4 with i.i.d. copies $(X_1, Y_1), \dots, (X_n, Y_n), (\tilde{X}_n, \tilde{Y}_n)$ of (X, Y).

a) The left-hand side is bounded by

$$
\begin{aligned}
&\frac{n}{2}E\Bigg(\sum_{i=1}^{n}\int\frac{Y_i^{[n]}T_h(x-X_i)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}}K_h(x-X_j)}\mu(dx)\\
&\qquad-\sum_{i=1}^{n-1}\int\frac{Y_i^{[n]}T_h(x-X_i)}{1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)+K_h(x-\tilde{X}_n)}\mu(dx)\\
&\qquad-\int\frac{\tilde{Y}_n^{[n]}T_h(x-\tilde{X}_n)}{1+\sum_{j=1}^{n-1}K_h(x-X_j)}\mu(dx)\Bigg)^2\\
&\quad\le nE\left(\int\frac{Y_n^{[n]}T_h(x-X_n)+\tilde{Y}_n^{[n]}T_h(x-\tilde{X}_n)}{1+\sum_{j=1}^{n-1}K_h(x-X_j)}\mu(dx)\right)^2\\
&\qquad+nE\left(\sum_{i=1}^{n-1}\int\frac{Y_i^{[n]}T_h(x-X_i)[K_h(x-\tilde{X}_n)+K_h(x-X_n)]}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\right)^2\\
&\quad\le 4nE\left\{(Y_n^{[n]})^2\left(\int\frac{T_h(x-X_n)}{1+\sum_{j=1}^{n-1}K_h(x-X_j)}\mu(dx)\right)^2\right\}\\
&\qquad+4nE\left(\sum_{i=1}^{n-1}\int\frac{Y_i^{[n]}T_h(x-X_i)K_h(x-X_n)}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\right)^2\ \le\ \frac{c}{n}E(Y_n^{[n]})^2
\end{aligned}
$$

for some $c\in\mathbb{R}_+$, the latter by Lemma 3.8 a) (via conditioning with respect to X_n) and by Lemma 3.9 b).

b) The left-hand side is bounded by

$$
\begin{aligned}
&\frac{n}{2}E\Bigg(\sum_{i=1}^{n}\int\frac{Y_i^{[n]}T_h(x-X_i)EK_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}}K_h(x-X_j)]^2}\mu(dx)\\
&\qquad-\sum_{i=1}^{n-1}\int\frac{Y_i^{[n]}T_h(x-X_i)EK_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)+K_h(x-\tilde{X}_n)]^2}\mu(dx)\\
&\qquad-\int\frac{Y_n^{[n]}T_h(x-\tilde{X}_n)EK_h(x-X_{n+1})}{[1+\sum_{j=1}^{n-1}K_h(x-X_j)]^2}\mu(dx)\Bigg)^2\\
&\quad\le nE\left(\int\frac{Y_n^{[n]}T_h(x-X_n)+\tilde{Y}_n^{[n]}T_h(x-\tilde{X}_n)}{[1+\sum_{j=1}^{n-1}K_h(x-X_j)]^2}EK_h(x-X_{n+1})\mu(dx)\right)^2\\
&\qquad+2nE\Bigg(\sum_{i=1}^{n-1}\int\frac{Y_i^{[n]}T_h(x-X_i)[K_h(x-X_n)+K_h(x-\tilde{X}_n)]}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^3}\\
&\qquad\qquad\qquad\qquad\cdot EK_h(x-X_{n+1})\mu(dx)\Bigg)^2\\
&\quad\le 4nE\left\{(Y_n^{[n]})^2\left(\int\frac{T_h(x-X_n)EK_h(x-X_{n+1})}{[1+\sum_{j=1}^{n-1}K_h(x-X_j)]^2}\mu(dx)\right)^2\right\}
\end{aligned}
$$

$$+ 8nE\left(\sum_{i=1}^{n-1}\int \frac{Y_i^{[n]}T_h(x-X_i)K_h(x-X_n)EK_h(x-X_{n+1})}{[1+\sum_{j\in\{1,\dots,n-1\}\setminus\{i\}}K_h(x-X_j)]^3}\mu(dx)\right)^2$$
$$\leq \frac{c}{n^3}E(Y_n^{[n]})^2$$

for some $c \in \mathbb{R}_+$, the latter by Lemma 3.8 d) (via conditioning with respect to X_n) and by Lemma 3.9 c).

c) Immediately by Lemma 3.9 a),b). $\square$

Set
$$m'_n(x) := \sum_{i=1}^{n}\frac{Y_iK_{h_n}(x-X_i)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{i\}}K_{h_n}(x-X_j)}, \qquad x \in \mathbb{R}^d.$$

LEMMA 3.11. *Let K be a regular kernel and let (h_n) satisfy $0 < h_n \to 0$, $nh_n^d \to \infty$.*

a) *If Y is bounded, then*

$$\int\left[\frac{\sum_{i=1}^n Y_iK_{h_n}(x-X_i)}{\sum_{i=1}^n K_{h_n}(x-X_i)} - \frac{\sum_{i=1}^n Y_iK_{h_n}(x-X_i)}{1+\sum_{i=1}^n K_{h_n}(x-X_i)}\right]\mu(dx) \to 0 \quad a.s., \tag{3.7}$$

$$\int |\widehat{m}_n(x) - m_n(x)|\mu(dx) \to 0 \quad a.s., \tag{3.8}$$

$$\int |m'_n(x) - m_n(x)|\mu(dx) \to 0 \quad a.s. \tag{3.9}$$

b) *For each $l \in \mathbb{N}$*

$$\int \frac{K_{h_n}(x-X_l)}{1+\sum_{j\in\{1,\dots,n\}\setminus\{l\}}K_{h_n}(x-X_j)}\mu(dx) \to 0 \quad a.s.$$

c) *If $EY < \infty$, then*

$$E\int m'_n(x)\mu(dx) \to EY. \tag{3.10}$$

PROOF. a) Let $Y_i \leq L$, $i \in \mathbb{N}$, for some $L \in \mathbb{R}_+$. An upper bound for the left-hand side of (3.7) is
$$L\int \frac{1}{1+\sum_{i=1}^n K_{h_n}(x-X_i)}\mu(dx).$$

We have
$$\left|\int \frac{1}{1+\sum_{i=1}^n K_{h_n}(x-X_i)}\mu(dx) - \int \frac{1}{1+E\sum_{i=1}^n K_{h_n}(x-X_i)}\mu(dx)\right|$$
$$\leq \int\left|\frac{\sum_{i=1}^n K_{h_n}(x-X_i)}{E\sum_{i=1}^n K_{h_n}(x-X_i)} - 1\right|\mu(dx) \to 0 \quad \text{a.s.}$$

according to Devroye and Krzyżak (1989), especially p. 76. Further
$$\int \frac{1}{1+E\sum_{i=1}^n K_{h_n}(x-X_i)}\mu(dx) \to 0,$$

since $E\sum_{i=1}^n K_{h_n}(x-X_i) \to \infty$ for μ-almost all x, because of $nh_n^d \to \infty$ and

$$\liminf \frac{\int K_{h_n}(x-t)\mu(dt)}{h_n^d} \geq b \liminf \frac{\mu(x+S_{0,rh_n})}{h_n^d} > 0$$

for μ-almost all x. For the latter relation see the proof of Lemma 2.2 in Devroye (1981), compare also Lemma 2 in Greblicki *et al.* (1984) and Lemma 24.6 in Györfi *et al.* (2002). Thus (3.7) is obtained. From this we obtain (3.8) and (3.9) because of

$$\frac{\sum_{i=1}^n Y_i K_{h_n}(x-X_i)}{1+\sum_{i=1}^n K_{h_n}(x-X_i)} \leq \left\{ \begin{matrix} m_n(x) \\ m_n'(x) \end{matrix} \right\} \leq \widehat{m}_n(x),$$

which holds for m_n because of $0 < \delta \leq 1$ and for m_n' because of $0 \leq K \leq 1$.

b) The assertion, even for general $h_n > 0$, follows from

$$\sum_{n=1}^{\infty} E\left[\int \frac{K_{h_n}(x-X_l)}{1+\sum_{j\in\{1,\ldots,n\}\setminus\{l\}} K_{h_n}(x-X_j)}\mu(dx)\right]^2 < \infty,$$

which is obtained from Lemma 3.8 a) (with $T = K$) via conditioning.

c) In the case of bounded Y we have

$$\int \widehat{m}_n(x)\mu(dx) \to \int m(x)\mu(dx) = EY \qquad \text{a.s.}$$

according to Lemma 3.7, and then obtain (3.10) by (3.8), (3.9) and the dominated convergence theorem. For general integrable $Y \geq 0$ we have

$$\begin{aligned} E\int m_n'(x)\mu(dx) &= n\int E\frac{Y_1 K_{h_n}(x-X_1)}{1+\sum_{j=2}^n K_{h_n}(x-X_j)}\mu(dx) \\ &\leq cEY, \qquad n \in \mathbb{N}, \end{aligned}$$

for some $c \in \mathbb{R}_+$ by Lemma 3.8 a) (with $T = K$) via conditioning. From these results we obtain (3.10) for integrable Y by Lemma 3.1 b). □

LEMMA 3.12. *Let K be a smooth kernel and (h_n) as in Theorem 2.1. Then a regular kernel $\tilde{M}$ exists such that*

$$K_{h_n} - K_{h_{n+1}} \leq \frac{1}{n}\tilde{M}_{h_n} := \frac{1}{n}\tilde{M}\left(\frac{\cdot}{h_n}\right), \qquad n \in \mathbb{N},$$

and that $\tilde{M}^2/K$ (with $0/0 := 0$) is also a regular kernel.

PROOF. We use the notation of Definition 2.1 and notice that

$$s|H'(s)| = \sqrt{R(s)}\sqrt{H(s)} \leq \frac{1}{2}R(s) + \frac{1}{2}H(s).$$

The function $s \to s|H'(s)|$, $s \in \mathbb{R}_+$ is continuous and, for s sufficiently large, nonincreasing, further $\int s|H'(s)|s^{d-1}ds < \infty$. Choose $b \in (1,\infty)$, $c > 0$ such that

$$\frac{h_n}{h_{n+1}} \leq b, \qquad \frac{1}{h_{n+1}} - \frac{1}{h_n} \leq \frac{1}{h_n}\frac{c}{n}$$

for all n, and set

$$M(s) := c \sup_{s \le v \le bs} v|H'(v)|, \qquad Q(s) := M^2(s)/H(s), \qquad s \ge 0 \ (0/0 := 0).$$

M is continuous. For u sufficiently large, M satisfies $M(u) = cu|H'(u)|$ and is nonincreasing. Further $\int M(s)s^{d-1}ds < \infty$, i.e., $M(\|\cdot\|)$ is integrable. Q is piecewise continuous. For u sufficiently large, Q satisfies $Q(u) = c^2R(u)$ and is nonincreasing. Also $\int Q(s)s^{d-1}ds < \infty$, i.e., $Q(\|\cdot\|)$ is integrable. Set $\tilde{M}(\cdot) := M(\|\cdot\|)$, $\tilde{Q}(\cdot) := Q(\|\cdot\|)$. $\tilde{M}$ and $\tilde{Q} = \tilde{M}^2/K$ are regular kernels. Now

$$\begin{aligned} n\left[K\left(\frac{t}{h_n}\right) - K\left(\frac{t}{h_{n+1}}\right)\right] &\le c \sup_{\|t\|/h_n \le s \le \|t\|/h_{n+1}} |H'(s)| \cdot \frac{\|t\|}{h_n} \\ &\le M\left(\frac{\|t\|}{h_n}\right) = \tilde{M}_{h_n}(t), \qquad t \in \mathbb{R}^d. \end{aligned}$$ □

LEMMA 3.13. *Let K be a smooth kernel and (h_n) as in Theorem 2.1. Assume $EY < \infty$. Then*

$$\int m_n'(x)\mu(dx) \to EY \qquad a.s. \tag{3.11}$$

PROOF. Set

$$\begin{aligned} m_n^{(n)}(x) &:= \sum_{i=1}^n \frac{Y_i^{[n]} K_{h_n}(x - X_i)}{1 + \sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_{h_n}(x - X_j)}, \\ m_n^*(x) &:= \sum_{i=1}^n \frac{Y_i^{[i]} K_{h_n}(x - X_i)}{1 + \sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_{h_n}(x - X_j)}, \qquad x \in \mathbb{R}^d. \end{aligned}$$

We notice

$$m_n^* \le m_n^{(n)} \le m_n', \qquad n \in \mathbb{N}. \tag{3.12}$$

Since a.s. $Y_n = Y_n^{[n]}$ from some random index on (by Lemma 3.5 a) because of $EY < \infty$), we have

$$\int (m_n'(x) - m_n^*(x))\mu(dx) \to 0 \qquad \text{a.s.} \tag{3.13}$$

by Lemma 3.11 b).

In the first step we show

$$\frac{1}{N}\sum_{n=1}^N \int m_n^{(n)}(x)\mu(dx) \to EY \qquad \text{a.s.} \tag{3.14}$$

and thus, by (3.12) and (3.13),

$$\frac{1}{N}\sum_{n=1}^N \int m_n^*(x)\mu(dx) \to EY \qquad \text{a.s.} \tag{3.15}$$

We have

$$\sum \frac{E[\int m_n^{(n)}(x)\mu(dx) - E\int m_n^{(n)}(x)\mu(dx)]^2}{n} \le c \sum \frac{E(Y_n^{[n]})^2}{n^2}$$

$$\text{(by Lemma 3.10 a) with } T = K)$$

$$< \infty \quad \text{(by } EY < \infty \text{ and Lemma 3.5 b)).}$$

Thus

$$\sum \frac{[\int m_n^{(n)}(x)\mu(dx) - E\int m_n^{(n)}(x)\mu(dx)]^2}{n} < \infty \quad \text{a.s.,}$$

and the Kronecker lemma together with the Cauchy-Schwarz inequality yields

$$\frac{1}{N}\sum_{n=1}^{N}\left(\int m_n^{(n)}(x)\mu(dx) - E\int m_n^{(n)}(x)\mu(dx)\right) \to 0 \quad \text{a.s.} \tag{3.16}$$

With $Y^{(L)} := YI_{[Y\le L]}$, $L > 0$, we notice

$$\begin{aligned}
&\limsup E\int [m_n'(x) - m_n^{(n)}(x)]\mu(dx)\\
&\le \lim E\int \sum_{i=1}^{n} \frac{(Y_i - Y_i^{(L)})K_{h_n}(x - X_i)}{1 + \sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_{h_n}(x - X_j)}\mu(dx)\\
&= E(Y - Y^{[L]}) \to 0 \quad (L \to \infty),
\end{aligned}$$

where the equation follows from Lemma 3.11 c) for $Y_i - Y_i^{[L]}$ instead of Y_i. Then once more by Lemma 3.11 c), we obtain

$$E\int m_n^{(n)}(x)\mu(dx) \to EY. \tag{3.17}$$

(3.16) and (3.17) yield (3.14).

In the second step we show

$$\int m_n^{(*)}(x)\mu(dx) \to EY \quad \text{a.s.} \tag{3.18}$$

by use of (3.15). We apply Lemma 3.3. Noticing $Y_i \ge 0$, $K \ge 0$, $K_{h_n} \ge K_{h_{n+1}}$ (by $H \downarrow$, $h_n \downarrow$) and $Y_i^{[i]} \le Y_i^{[n]}$ $(i = 1, \dots, n)$, $K_{h_n} - K_{h_{n+1}} \le \frac{1}{n}\tilde{M}_{h_n}$ (according to Lemma 3.12), we obtain

$$\int m_{n+1}^*(x)\mu(dx) - \int m_n^*(x)\mu(dx) \ge -C_n - D_n, \qquad n \in \mathbb{N},$$

with

$$\begin{aligned}
C_n &= \frac{1}{n}\int \sum_{i=1}^{n} \frac{Y_i^{[n]}\tilde{M}_{h_n}(x - X_i)}{1 + \sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_{h_n}(x - X_j)}\mu(dx),\\
D_n &= \int \sum_{i=1}^{n} \frac{Y_i^{[n]}K_{h_{n+1}}(x - X_i)K_{h_{n+1}}(x - X_{n+1})}{[1 + \sum_{j\in\{1,\dots,n\}\setminus\{i\}} K_{h_{n+1}}(x - X_j)]^2}\mu(dx).
\end{aligned}$$

Setting

$$F_n := \int \sum_{i=1}^{n} \frac{Y_i^{[n]} K_{h_{n+1}}(x - X_i)(K_{h_{n+1}}(x - X_{n+1}) - EK_{h_{n+1}}(x - X_{n+1}))}{[1 + \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} K_{h_{n+1}}(x - X_j)]^2} \mu(dx),$$

$$G_n := \int \sum_{i=1}^{n} \frac{Y_i^{[n]} K_{h_{n+1}}(x - X_i)}{[1 + \sum_{j\in\{1,\ldots,n\}\setminus\{i\}} K_{h_{n+1}}(x - X_j)]^2} EK_{h_{n+1}}(x - X_{n+1})\mu(dx)$$

we have $D_n = F_n + G_n$ and thus the representation

$$\begin{aligned} \int m_{n+1}^*(x)\mu(dx) &- \int m_n^*(x)\mu(dx) \\ &= J_n - C_n - F_n - G_n \\ &= J_n - (EC_n + EG_n) - (C_n - EC_n) - (G_n - EG_n) - F_n \end{aligned} \tag{3.19}$$

where

$$J_n \geq 0. \tag{3.20}$$

Via conditioning we have

$$\begin{aligned} EC_n &\leq \int E(Y_1 \mid X_1 = z) \left[E \int \frac{\tilde{M}_{h_n}(x - z)}{1 + \sum_{j\in\{2,\ldots,n\}} K_{h_n}(x - X_j)} \mu(dx) \right] \mu(dz) \\ &\leq \frac{c}{n} \int E(Y_1 \mid X_1 = z)\mu(dz) \qquad \text{(by Lemma 3.8 a) for the regular kernel } T = \tilde{M}) \\ &= \frac{c}{n} EY, \\ EG_n &\leq n \int E(Y_1 \mid X_1 = z) \left[E \int \frac{K_{h_{n+1}}(x - z) EK_{h_{n+1}}(x - X_{n+1})}{[1 + \sum_{j\in\{2,\ldots,n\}} K_{h_{n+1}}(x - X_j)]^2} \mu(dx) \right] \mu(dz) \\ &\leq \frac{c}{n} \int E(Y_1 \mid X_1 = z)\mu(dz) \qquad \text{(by Lemma 3.8 d) with } T = K) \\ &= \frac{c}{n} EY \end{aligned}$$

for some $c \in \mathbb{R}_+$. Thus for $N > M \to \infty$ with $N/M \to 1$ we obtain

$$\sum_{n=M}^{N} (EC_n + EG_n) \to 0. \tag{3.21}$$

Noticing that $\tilde{M}$ and $\tilde{M}^2/K = \tilde{Q}$ are regular kernels, we have $\operatorname{Var}(C_n) \leq \frac{c}{n^3} E(Y_n^{[n]})^2$ by Lemma 3.10 a) with $T = \tilde{M}$. Further $\operatorname{Var}(G_n) \leq \frac{c}{n^3} E(Y_n^{[n]})^2$ by Lemma 3.10 b) with $T = K$. Thus, by Lemma 3.5 b), $\sum n \operatorname{Var}(C_n) < \infty$, $\sum n \operatorname{Var}(G_n) < \infty$, which yields

$$\sum n[(C_n - EC_n) + (G_n - EG_n)]^2 < \infty \quad \text{a.s.} \tag{3.22}$$

Finally, by Lemma 3.10 c) with $T = K$, and Lemma 3.5 b), we have $\sum EF_n^2 \leq c \sum \frac{1}{n^2} E(Y^{[n]})^2 < \infty$, which yields

$$\text{a.s. convergence of } \sum F_n, \tag{3.23}$$

because (F_n) is a martingale difference sequence. Now from (3.15) together with (3.19), (3.20), (3.21), (3.22), (3.23) we obtain (3.18) by Lemma 3.3.

Finally (3.18) and (3.13) yield the assertion (3.11). □

PROOF OF THEOREM 2.1. In the case of bounded $Y \leq L$ we have $\int |\widehat{m}_n(x) - m(x)|\mu(dx) \to 0$ a.s. according to Lemma 3.7, and then obtain

$$\int |m_n(x) - m(x)|^2 \mu(dx) \leq L \int |m_n(x) - m(x)|\mu(dx) \to 0 \quad \text{a.s.} \tag{3.24}$$

by Lemma 3.11 a). For general integrable $Y \geq 0$ we notice

$$m_n \leq \frac{2}{\delta} m'_n \tag{3.25}$$

(by $\max\{\delta, t\} \geq (\delta + t)/2$ for $t \geq 0$), apply Lemma 3.13 and obtain

$$\limsup \int m_n(x)\mu(dx) \leq \frac{2}{\delta} \lim \int m'_n(x)\mu(dx) = \frac{2}{\delta} EY \quad \text{a.s.} \tag{3.26}$$

From (3.24) and (3.26) we obtain (1.1) for square integrable Y by Lemma 3.1 a). □

Acknowledgements

The author would like to thank the referee for valuable suggestions which improved the readability of the paper.

REFERENCES

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

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

Devroye, L. and Wagner, T. J. (1980*b*). On the L_1 convergence of kernel estimators of regression functions with applications in discrimination, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **51**, 15–25.

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

Efron, B. and Stein, C. (1981). The jackknife estimate of variance, *Annals of Statistics*, **9**, 586–596.

Greblicki, W., Krzyżak, A. and 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, *Nonparametric Functional Estimation and Related Topics* (ed. G. Roussas), 329–338, NATO ASI Ser. C., Kluwer Academic Publishers, Dordrecht.

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

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

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

Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-free Theory of Nonparametric Regression*, Springer-Verlag, New York, Berlin.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, Massachusetts.

Hardy, G. H. (1949). *Divergent Series*, Oxford University Press, Oxford.

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*, **68**, 138–164.

Kozek, A. S., Leslie, J. R. and Schuster, E. F. (1998). On a universal strong law of large numbers for conditional expectations, *Bernoulli*, **4**, 143–165.

Krzyżak, A. and Pawlak, M. (1987). The pointwise rate of convergence of the kernel regression estimate, *Journal of Statistical Planning and Inference*, **16**, 159–166.

Loève, M. (1977). *Probability Theory*, 4th ed., Springer-Verlag, New York, Berlin.

Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization, *IEEE Transactions on Information Theory*, **41**, 677–687.

Nadaraya, E. A. (1964). On estimating regression, *Theory of Probability and Its Applications*, **9**, 141–142.

Pawlak, M. (1991). On the almost everywhere properties of the kernel regression estimate, *Annals of the Institute of Statistical Mathematics*, **43**, 311–326.

Révész, P. (1973). Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I, *Studia Scientiarum Mathematicarum Hungarica*, **8**, 391–398.

Spiegelman, C. and Sacks, J. (1980). Consistent window estimation in nonparametric regression, *Annals of Statistics*, **8**, 240–246.

Steele, J. M. (1986). An Efron-Stein inequality for nonsymmetric statistics, *Annals of Statistics*, **14**, 753–758.

Stone, C. J. (1977). Consistent nonparametric regression, *Annals of Statistics*, **5**, 595–645.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators, *Annals of Statistics*, **8**, 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *Annals of Statistics*, **10**, 1040–1053.

Stute, W. (1986). On almost sure convergence of conditional empirical distribution functions, *Annals of Probability*, **14**, 891–901.

Walk, H. (2001). Strong universal pointwise consistency of recursive regression estimates, *Annals of the Institute of Statistical Mathematics*, **53**, 691–707.

Walk, H. (2005). Strong laws of large numbers by elementary Tauberian arguments, *Monatshefte für Mathematik*, **144**, 329–346.

Watson, G. S. (1964). Smooth regression analysis, *Sankhyā Series A*, **26**, 359–372.

Zeller, K. and Beekmann, W. (1970). *Theorie der Limitierungsverfahren*, 2. Aufl., Springer-Verlag, Berlin.