Ann. Inst. Statist. Math.
Vol. 53, No. 4, 691–707 (2001)

STRONG UNIVERSAL POINTWISE CONSISTENCY OF RECURSIVE REGRESSION ESTIMATES

HARRO WALK

Mathematisches Institut A, Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany

(Received January 7, 1999; revised February 28, 2000)

Abstract. For semi-recursive and recursive kernel estimates of a regression of Y on X (d-dimensional random vector X, integrable real random variable Y), introduced by Devroye and Wagner and by Révész, respectively, strong universal pointwise consistency is shown, i.e. strong consistency P_X-almost everywhere for general distribution of (X, Y). Similar results are shown for the corresponding partitioning estimates.

Key words and phrases: Nonparametric regression estimation, semi-recursive estimation, recursive estimation, kernel estimates, partitioning estimates, strong universal pointwise consistency, strong laws of large numbers, conditional expectations, truncation, covering.

1. Introduction

Let X be a d-dimensional random vector with distribution $P_X =: \mu$ and let Y be a real random variable with $E|Y| < \infty$. The regression function $m : \mathbb{R}^d \to \mathbb{R}$ defined by $m(x) := E(Y \mid X = x)$, shall be estimated on the basis of an observable training sequence (X_1, Y_1), $(X_2, Y_2), \ldots$ of independent copies of (X, Y) where no further assumption on the distribution of (X, Y) is used. Let in the context of observations $(x_1, y_1), \ldots, (x_n, y_n)$ of $(X_1, Y_1), \ldots, (X_n, Y_n)$ a nonparametric estimator of $m(x)$ be denoted by $m_n(x_1, y_1, \ldots, x_n, y_n; x) =: m_n(x)$, $x \in \mathbb{R}^d$. To simplify the notation the abbreviation mod μ is used to indicate that a relation holds for μ-almost all $x \in \mathbb{R}^d$. I denotes an indicator function, S denotes a closed sphere in $\mathbb{R}^d$ around 0 with finite positive radius.

The estimation sequence (m_n) is called *strongly* [*weakly*] *universally pointwise consistent*, if

$$\text{almost surely (a.s.) } m_n \to m(x) \text{ mod } \mu$$
$$[m_n(x) \to m(x) \text{ in probability mod } \mu]$$

for all distributions of (X, Y) with $E|Y| < \infty$. We shall use this definition although in literature often the last condition is replaced by the stronger condition $EY^2 < \infty$. In the case that, with $EY^2 < \infty$, pointwise consistency is replaced by consistency in $L^2(\mu)$ sense, one speaks of strong [weak] universal consistency.

Stone (1977) first pointed out that there exist weak universal consistent estimations.

The aim of this paper is to show strong universal pointwise consistency for some (wide sense) recursive estimation sequences, namely semi-recursive kernel and partitioning estimates and (narrow sense) recursive kernel and partitioning estimates with suitable kernels and bandwidth sequences and suitable sequences of partitions, respectively.

For a sequence of measurable functions $K_n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+ (n \in \mathbb{N})$, we first consider estimates of the form

$$m_n(x) := \frac{\sum_{i=1}^n Y_i K_i(x, X_i)}{\sum_{i=1}^n K_i(x, X_i)} \qquad \left(\frac{0}{0} := 0\right). \tag{1.1}$$

Such an estimate is called semi-recursive because it can be updated sequentially by adding extra terms to both the numerator and the denominator when new observations become available.

The semi-recursive kernel estimate is defined by a kernel $K : \mathbb{R}^d \to \mathbb{R}_+$ and a sequence of bandwidths $h_n > 0$ via (1.1) with

$$K_n(x, u) := K\left(\frac{x-u}{h_n}\right). \tag{1.2}$$

This estimate was introduced by Devroye and Wagner (1980*b*), who investigated, as Krzyżak (1992), L_1-convergence. Its definition is motivated by a recursive density estimate due to Wolverton and Wagner (1969) and Yamato (1971), compare also Greblicki's (1974) and Ahmad and Lin's (1976) regression estimate. Krzyżak and Pawlak (1984) and Greblicki and Pawlak (1987) showed weak universal pointwise consistency, also strong pointwise consistency in the case $E|Y|^p < \infty$ for some $p > 1$, under some conditions on the kernel and the bandwidth sequence. Györfi *et al.* (1998) treated weak and strong universal consistency.

In the non-recursive kernel estimate (Nadaraya (1964), Watson (1964)), $K_i (i = 1, \ldots, n)$ in (1.1) is replaced by K_n in (1.2). Its weak universal consistency for rather general kernels and bandwidth sequences was proved by Devroye and Wagner (1980*a*) and Spiegelman and Sacks (1980). Strong universal consistency for special bandwidth sequences was proved by Walk (2000). Weak universal pointwise consistency for suitable kernels was shown by Devroye (1981) and Greblicki *et al.* (1984). Whether strong universal pointwise consistency holds, is an open problem. Kozek *et al.* (1998) showed that for window kernel $K = I_S$ strong pointwise consistency holds under a slightly sharpened integrability assumption on Y, see also Stute (1986), or an assumption κ on μ fulfilled e.g. for absolutely continuous μ. The integrability condition means $E\Phi(|Y|) < \infty$ for some symmetric convex $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\Phi'(2t) < C\Phi'(t)(0 < C < \infty)$, $\Phi(0) = 0$, $\Phi(\sqrt{\cdot})$ subadditive, which is used together with the condition $h_n \downarrow 0$, $\sum h_n^d \Phi(nh_n^d)^{-1} < \infty$ concerning the bandwidths; e.g. $\Phi(t) = t(c^* + \ln(1+t)(\ln\ln(1+t))^r)$ $(r > 1$, suitable $c^* > 0)$ and $h_n = cn^{-\gamma}$ $(0 < c < \infty, 0 < \gamma d < 1)$. The assumption κ means

$$0 < \liminf \mu(x + h_n S)n^{1-\kappa} \leq \limsup \mu(x + h_n S)n^{1-\kappa} < \infty$$

μ-almost everywhere for some $\kappa \in [1 - \delta d, 1]$ which can also be replaced by countably many $\kappa_i(x) \in [1 - \gamma d, 1]$, where $h_n = cn^{-\gamma}(0 < c < \infty, 0 < \gamma d < 1)$. Greblicki and Pawlak (1985) used Dirichlet kernels and kernels associated with the Hermite series and showed weak and strong pointwise consistency for integrable and uniformly bounded Y, respectively, in the case that $X_1, X_2, \ldots$ are real random variables with a density.

The semi-recursive partitioning or histogram estimator is defined analogously via (1.1) in the context of a sequence of partitions $\mathcal{P}_n = \{A_{n1}, A_{n2}, \ldots\}$ (finite or denumerable family of Borel sets) of $\mathbb{R}^d$ where

$$K_n(x, u) := I_{A_n(x)}(u), \tag{1.3}$$

$A_n(x)$ denoting the set A_{ni} with $x \in A_{ni}$. We require that the sequence of partitions is nested, i.e. that the sequence of generated σ-algebras $\mathcal{F}(\mathcal{P}_n)$ increases. The estimator can then be computed efficiently by storing the constant denumerator and denominator for each cell of $(\mathcal{P}_n)$. Weak and strong universal consistency were studied by Györfi *et al.* (1998).

The non-recursive partitioning estimate, where K_i $(i = 1, \ldots, n)$ in (1.1) is replaced by K_n in (1.3) was investigated by Devroye and Györfi (1983), Györfi (1991) and Walk (2000). Also here the problem whether strong universal pointwise consistency holds, is open.

For the function sequence $(\bar{K}_n) := (h_n^{-d} K_n)$ with general K_n as before and positive numbers h_n we also consider recursive estimates m_n $(n \in \mathbb{N})$ of stochastic approximation type defined by

$$m_1 = Y_1, \tag{1.4}$$

$$m_{n+1}(x) = m_n(x) - a_{n+1} m_n(x) \bar{K}_{n+1}(x, X_{n+1}) + a_{n+1} Y_{n+1} \bar{K}_{n+1}(x, X_{n+1}) \tag{1.5}$$

with positive numbers a_n, so-called gains.

Equations (1.4), (1.5) with K_n given by (1.2) define the recursive kernel regression estimate introduced and investigated by Révész (1973). Strong universal consistency and strong pointwise consistency of this estimate was proven for suitable kernels, gains and bandwidths by Györfi and Walk (1997) under the assumption $EY^2 < \infty$.

Equations (1.4), (1.5) with K_n given by (1.3) define a recursive partitioning regression estimate.

In this paper we show strong universal pointwise consistency for semi-recursive kernel estimates with special kernels, mainly window kernels (Theorem 2.1a) and for semi-recursive partitioning estimates (Theorem 2.2) allowing rather general bandwidth sequences and (nested) partitioning sequences, respectively, in contrast to the non-recursive case. In the proofs we use truncation and the covering argument. Obviously the results can be considered as strong laws of large numbers for conditional expectations. In the case of more general kernels with compact support, strong pointwise consistency is established under condition $E|Y| \ln^+ |Y| < \infty$ (Theorem 2.1b). Considering a general structure of $\bar{K}_n$ (in a proposition) we prove strong universal pointwise consistency for recursive kernel estimates, with rather general kernel, and for recursive partitioning estimates (Theorems 2.3 and 2.4, resp.) under rather strong conditions on the gain sequence and the bandwidth and partitioning sequence, resp., which cannot be essentially weakened.

The results are presented in Section 2, the proofs are given in Section 3.

2. Results

The first result deals with the semi-recursive kernel estimator of Devroye and Wagner (1980*b*).

THEOREM 2.1. *For $n \in \mathbb{N}$ let*

$$m_n(x) := \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_i}\right)}, \qquad x \in \mathbb{R}^d,$$

with symmetric measurable $K : \mathbb{R}^d \to \mathbb{R}_+$ *and* $h_n \downarrow 0$, $\sum h_n^d = \infty$.

a) *If* $\alpha I_S \le K \le \beta I_S$ *for some* $0 < \alpha < \beta < \infty$, *then a.s.* $m_n(x) \to m(x) \bmod \mu$.

b) *If* $\alpha H(\|x\|) \le K(x) \le \beta H(\|x\|)$, $x \in \mathbb{R}^d$, *for some* $0 < \alpha < \beta < \infty$ *and nonincreasing* $H : \mathbb{R}_+ \to \mathbb{R}_+$ *with compact support and* $H(+0) > 0$ *and if additionally* $E|Y| \ln^+ |Y| < \infty$ *is assumed, then a.s.* $m_n(x) \to m(x) \bmod \mu$.

Remark 2.1. One can obtain Theorem 1 of Kozek *et al.* (1998) concerning the nonrecursive case with $K = I_S$ from Theorem 2.1a) for $h_n = cn^{-\gamma}$, $0 < c < \infty$, $0 < \gamma d < 1$, noticing

$$K\left(\frac{x - X_i}{h_n}\right) \le K\left(\frac{x - X_i}{h_i}\right) \quad \text{for} \quad i = 1, \dots, n$$

and

$$\text{a.s.} \qquad \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)}{n\mu(x + h_n S)} \to m(x) \bmod \mu$$

for bounded Y (Devroye (1981)), especially for $Y = 1$ and $m(x) = 1$, by use of assumption κ quoted above and of the argument in the proof of Lemma 3.1. Under the same assumptions a similar argument yields strong pointwise consistency of Greblicki's (1974) and Ahmad and Lin's (1976) kernel estimate for which one has in (1.1) $h_i^{-d} K((x-X_i)/h_i)$ instead of $K((x - X_i)/h_i)$.

Remark 2.2. The condition $\sum h_n^d = \infty$ is also necessary for both general assertions in Theorem 2.1 as the following argument shows. In the case of an independent family $\{X_1, Y_1, X_2, Y_2, \dots\}$ with X_i uniformly distributed on $[0,1]^d$ and $P[Y_i = 1] = P[Y_i = -1] = 1/2$, the assumption $\sum h_n^d < \infty$ would yield —for each $x \in [0,1]^d$— convergence of the series $\sum EK((x - X_i)/h_i)$, $\sum \operatorname{Var} K((x - X_i)/h_i)$ and $\sum \operatorname{Var}(Y_i K((x - X_i)/h_i))$ with real-positive series sums, say E, V and W, respectively, where $\sum E(Y_i K((x - X_i)/h_i)) = 0$, thus, according to Sections 16.3 and 17.2 in Loève (1977), a.s. convergence of $(\sum_{i=1}^n K((x - X_i)/h_i))$ to a real-positive random variable, further a.s. and quadratic mean convergence of $(\sum_{i=1}^n Y_i K((x - X_i)/h_i))$ to a random variable with expectation 0 and variance W, therefore a.s. convergence of $(m_n(x))$ to a real random variable not degenerate to $0 = m(x)$.

The following result on semi-recursive partitioning estimates is an analogue to Theorem 1a. λ denotes Lebesgue measure.

THEOREM 2.2. *Let* $(\mathcal{P}_n)_{n\in\mathbb{N}}$ *be a nested sequence of partitions* $\mathcal{P}_n = \{A_{n1}, A_{n2}, \dots\}$ *of* $\mathbb{R}^d$ *by Borel sets such that*

$$\operatorname{diam} A_n(z) := \sup_{u,v \in A_n(z)} \|u - v\| \to 0 \ (n \to \infty), \qquad \sum \lambda(A_n(z)) = \infty$$

for each $z \in \mathbb{R}^d$. *For* $n \in \mathbb{N}$ *let*

$$m_n(x) := \frac{\sum_{i=1}^n Y_i I_{A_i(x)}(X_i)}{\sum_{i=1}^n I_{A_i(x)}(X_i)}, \qquad x \in \mathbb{R}^d.$$

Then a.s. $m_n(x) \to m(x) \bmod \mu$.

Now recursive estimates of stochastic approximation type are considered. Theorems 2.3 and 2.4 on the kernel method of Révész (1973) and its partitioning analogue, resp., are consequences of the following proposition which will be stated first.

PROPOSITION. *Let $(\bar{K}_n)_{n\geq 2}$ be a sequence of measurable nonnegative functions on $\mathbb{R}^d \times \mathbb{R}^d$ such that for every distribution μ of X*

$$\frac{\int \bar{K}_n(x,z)f(z)\mu(dz)}{\int \bar{K}_n(x,z)\mu(dz)} \to f(x) \bmod \mu \tag{2.1}$$

for all μ-integrable functions f on $\mathbb{R}^d$ and

$$\liminf_n \int \bar{K}_n(x,z)\mu(dz) > 0 \bmod \mu. \tag{2.2}$$

Let $(a_n)_{n\geq 2}$ be a sequence of positive numbers such that

$$\sum a_n = \infty, \tag{2.3}$$

$$1 > a_n \sup_{x,z} \bar{K}_n(x,z) = O\left(\frac{1}{n}\right). \tag{2.4}$$

If the sequence $(m_n)_{n\in\mathbb{N}}$ is defined by (1.4), (1.5), then a.s. $m_n(x) \to m(x) \bmod \mu$.

THEOREM 2.3. *Let K be a symmetric measurable nonnegative function on $\mathbb{R}^d$ satisfying $\alpha H(\|x\|) \leq K(x) \leq \beta H(\|x\|), x \in \mathbb{R}^d$, for some $0 < \alpha < \beta < \infty$ and a nonincreasing $H : \mathbb{R} \to \mathbb{R}_+$ with $H(+0) > 0$ and $r^d H(r) \to 0 (r \to \infty)$. Let $(h_n)_{n\geq 2}$ and $(a_n)_{n\geq 2}$ be sequences of positive numbers satisfying*

$$h_n \to 0, \quad \sum a_n = \infty, \quad \sup_x K(x) \sup_n a_n/h_n^d < 1, \;\; a_n/h_n^d = O(1/n)$$

(e.g. $a_n = \frac{1}{n \log n}$, $h_n = \frac{1}{(\log n)^{1/d}}$ $(n = 2, 3, \ldots)$ with $K \leq 1$). If the sequence $(m_n)_{n\in\mathbb{N}}$ is defined by

$$m_1 = Y_1,$$

$$m_{n+1}(x) = m_n(x) - a_{n+1}m_n(x)\frac{1}{h_{n+1}^d}K\left(\frac{x - X_{n+1}}{h_{n+1}}\right)$$
$$+a_{n+1}Y_{n+1}\frac{1}{h_{n+1}^d}K\left(\frac{x - X_{n+1}}{h_{n+1}}\right),$$

then a.s. $m_n(x) \to m(x) \bmod \mu$.

Remark 2.3. The rather restrictive condition on (a_n) and (h_n) in Theorem 2.3 imposed in view of pointwise convergence under first moment condition cannot be essentially weakened as the following simple counterexample shows.

Counterexample. Let $d = 1$, μ be concentrated on $x^* \in \mathbb{R}$, $K = I_{[-1,1]}$, $a_n \downarrow 0$ and $h_n \downarrow 0$ such that $a_n/h_n \to 0$, $\rho_n := na_n/h_n \to \infty$, Y integrable, but with

$$\sum P[Y > n/\rho_n] = \infty. \tag{2.5}$$

Let (m_n) be defined as in Theorem 2.3. If a.s. $m_n(x^*) \to EY$, then

$$\text{a.s.} \qquad a_n Y_n \frac{1}{h_n} \leq a_n Y_n \frac{1}{h_n} K\left(\frac{x^* - X_n}{h_n}\right) \to 0,$$

thus by the Borel-Cantelli lemma

$$\sum P\left[Y_n > \frac{h_n}{a_n}\right] = \sum P\left[Y > \frac{h_n}{a_n}\right] < \infty$$

in contradiction to (2.5).

THEOREM 2.4. *Let $(\mathcal{P}_n)_{n\geq 2}$ be a nested sequence of partitions $\mathcal{P}_n = \{A_{n1}, A_{n2}, \ldots\}$ of $\mathbb{R}^d$ by Borel sets and let $(h_n)_{n\geq 2}$ and $(a_n)_{n\geq 2}$ be sequences of positive numbers such that*

$$h_n \to 0, \qquad \sum a_n = \infty, \quad 1 > a_n/h_n^d = O(1/n)$$
$$\operatorname{diam} A_n(z) \to 0 (n \to \infty), \qquad \liminf_n \frac{\lambda(A_n(z))}{h_n^d} > 0$$

for each $z \in \mathbb{R}^d$. If the sequence $(m_n)_{n\in\mathbb{N}}$ is defined by

$$m_1 = Y_1,$$

$$\begin{aligned} m_{n+1}(x) = \quad & m_n(x) - a_{n+1} m_n(x) \frac{1}{h_{n+1}^d} I_{A_{n+1}(x)}(X_{n+1}) \\ & + a_{n+1} Y_{n+1} \frac{1}{h_{n+1}^d} I_{A_{n+1}(x)}(X_{n+1}), \end{aligned}$$

then a.s. $m_n(x) \to m(x)$ mod μ.

Remark 2.4. In the case $d = 1$, the assumption in Theorems 2.2 and 2.4 that $(\mathcal{P}_n)$ is nested, can be cancelled, if each partition consists of non-accumulating intervals (see end of Section 3).

3. Proofs

Lemma 3.1 adapts Theorem 2 of Györfi (1991) for L^p-convergence to the case of pointwise convergence. Lemma 3.2 is a consequence of classical results of Abel, Dini, Pringsheim and Cesàro (compare Knopp (1956), Section 5.1). Both lemmas are used for the proofs of Theorems 2.1 and 2.2.

LEMMA 3.1. *Let (K_n) be a sequence of measurable nonnegative functions on $\mathbb{R}^d \times \mathbb{R}^d$ with $K_n(x, z) \leq K_{\max} \in \mathbb{R}_+$ for all $(x, z) \in \mathbb{R}^d \times \mathbb{R}^d$ and all $n \in \mathbb{N}$. Assume that for every distribution μ of X*

$$\frac{\int K_n(x, z) f(z) \mu(dz)}{\int K_n(x, z) \mu(dz)} \to f(x) \text{ mod } \mu \tag{3.1}$$

for all μ-integrable functions f on $\mathbb{R}^d$ and

$$\sum_n \int K_n(x,z)\mu(dz) = \infty \bmod \mu. \tag{3.2}$$

Assume further that a finite constant c^* exists such that

$$\text{a.s.} \quad \limsup_n \frac{\sum_{i=1}^n Y_i K_i(x, X_i)}{1 + \sum_{i=1}^n \int K_i(x,z)\mu(dz)} \leq c^* m(x) \bmod \mu \tag{3.3}$$

for all distributions $P_{(X,Y)}$ with $Y \geq 0$, $EY < \infty$. Let (m_n) be a sequence of estimates of the form

$$m_n(x) = \frac{\sum_{i=1}^n Y_i K_i(x, X_i)}{\sum_{i=1}^n K_i(x, X_i)}, \qquad x \in \mathbb{R}^d,$$

where $E \mid Y \mid < \infty$. Then a.s. $m_n(x) \to m(x) \bmod \mu$.

PROOF. Lemma 1 in Györfi *et al.* (1998) states the assertion under the assumptions (3.1) and (3.2) for (K_n) in the case of square integrable Y.

In the case of integrable Y which is assumed nonnegative without loss of generality, we use a truncation argument according to the proof of Theorem 2 in Györfi (1991). First we state

$$\text{a.s.} \quad \frac{\sum_{i=1}^n K_i(x, X_i)}{\sum_{i=1}^n \int K_i(x,z)\mu(dz)} \to 1 \bmod \mu$$

as in the proof of Lemma 1 in Györfi *et al.* (1998). This together with (3.2) and (3.3) yields

$$\text{a.s.} \quad \limsup_n m_n(x) \leq c^* m(x) \bmod \mu. \tag{3.4}$$

Now fix $\varepsilon > 0$. For all $L \in \mathbb{N}$, define $Y_j^L := Y_j I_{[Y_j \leq L]} + L I_{[Y_j > L]}$ and let m_L and m_{nL} be the functions m and m_n when (Y_j) is replaced by (Y_j^L). Then

$$\text{a.s.} \quad m_{nL}(x) - m_L(x) \to 0 \ (n \to \infty) \quad \text{for all} \quad L \in \mathbb{N} \bmod \mu$$

(see above) and with suitable $L_1(x) \in \mathbb{N}$

$$|m_L(x) - m(x)| < \varepsilon \quad \text{for all} \quad L \geq L_1(x) \bmod \mu,$$

further by (3.4)

$$\begin{aligned}\text{a.s.} \quad \limsup_n |m_n(x) - m_{nL}(x)| &\leq c^* E(Y_1 - Y_1^L \mid X_1 = x) \\ &\leq c^* \varepsilon \quad \text{for all} \quad L \geq L_1(x) \bmod \mu.\end{aligned}$$

These relations yield

$$\text{a.s.} \quad \limsup_n |m_n(x) - m(x)| \leq c^* \varepsilon + \varepsilon \bmod \mu$$

and thus the assertion. □

LEMMA 3.2. *Let* $0 \le r_n \le 1$; $R_n := r_1 + \cdots + r_n (n \in \mathbb{N})$, $R_0 := 0$. *There is a sequence* p_i *of integers with* $p_i \uparrow \infty$ *and*

$$R_{p_i} \le i + 1, \tag{3.5}$$

$$\sum_{j=p_i}^{\infty} \frac{r_j}{(1+R_j)^2} < \frac{1}{i}, \tag{3.6}$$

$$\sum_{j=1}^{p_i} \frac{r_j}{1+R_j} \le \ln(i+2) \ (i \in \mathbb{N}). \tag{3.7}$$

PROOF. Set $R_\infty := \lim R_n$ and $1/(1+R_\infty) := 0$ if $R_\infty = \infty$. For $p \in \{2, 3, \ldots\}$ we have

$$\sum_{j=p}^{\infty} \frac{r_j}{(1+R_j)^2} \le \sum_{j=p}^{\infty} \left(\frac{1}{1+R_{j-1}} - \frac{1}{1+R_j} \right) = \frac{1}{1+R_{p-1}} - \frac{1}{1+R_\infty}.$$

For $i \in \mathbb{N}$ choose $p_i \in \{2, 3, \ldots\}$ as the first index with

$$\frac{1}{1+R_{p_i-1}} - \frac{1}{1+R_\infty} < \frac{1}{i}.$$

Then (3.6) holds and by definition of p_i, $R_{p_i-2} \le i-1$, if $p_i \ge 3$, thus (3.5). Because of

$$\frac{r_j}{1+R_j} \le -\ln\left(1 - \frac{r_j}{1+R_j}\right), \quad j \in \mathbb{N},$$

and (3.5) we obtain

$$\sum_{j=1}^{p_i} \frac{r_j}{1+R_j} \le \ln(1+R_{p_i}) \le \ln(i+2), \quad i \in \mathbb{N},$$

i.e. (3.7). If the construction yields constant p_i from some index i^* on, then $r_j = 0$ for $j \ge p_{i^*}$, and we can replace p_i for $i > i^*$ by larger integers such that $p_i \uparrow \infty$. □

PROOF OF THEOREM 2.1. The proof consists of five steps. Only in the last two steps the special assumptions on K and Y in parts a) and b), resp., are distinguished. We use Lemma 3.1 with

$$K_n(x,t) = K\left(\frac{x-t}{h_n}\right).$$

In the first step notice that (3.1) and (3.2) follow from Lemma 1 of Greblicki *et al.* (1984) and (11) in Greblicki and Pawlak (1987).

It remains to verify (3.3) for $Y \ge 0$, $EY < \infty$. A covering argument and a truncation argument are used. Choose $R > 0$ such that $H(R) > 0$ (in part a) we choose R as the radius of S). Let the compact support of K be covered by finitely many closed spheres $S_1, \ldots, S_N$ each with radius $R/2$. Let $k \in \{1, \ldots, N\}$ be fixed. As in Györfi *et al.* ((1998), p. 13) for each $n \in \mathbb{N}$ and $t \in \mathbb{R}^d$ we show that $x \in t + h_n S_k$ implies

$$K\left(\frac{\cdot - x}{h_i}\right) \ge cK\left(\frac{\cdot - t}{h_i}\right) I_{S_k}\left(\frac{\cdot - t}{h_i}\right) \tag{3.8}$$

for all $i \in \{1, \ldots, n\}$ with $c = \alpha H(R)/\beta H(0) \in (0,1]$; compare Lemma 1 of Devroye and Krzyżak (1989). With $t = 0$ (without loss of generality) and $x/h_n = \tilde{x}$ it suffices to show $K(u - \tilde{x}h_n/h_i) \geq cK(u)$ for all $\tilde{x}$, $u \in S_k$ and all $n \in \mathbb{N}$, $i \in \{1, \ldots, n\}$. Because of

$$\left\| u - \tilde{x}\frac{h_n}{h_i} \right\| \leq \max_{0 \leq r \leq 1} \|u - r\tilde{x}\| = \max\{\|u\|, \|u - \tilde{x}\|\} \leq \max\{\|u\|, R\}$$

(since $h_n \leq h_i$) we have $H(\|u - \tilde{x}\frac{h_n}{h_i}\|) \geq H(R)H(\|u\|)/H(0)$ in both cases $\|u - \tilde{x}h_n/h_i\| \leq \|u\|$, $\|u - \tilde{x}h_n/h_i\| \leq R$ by monotonicity of H, and thus the desired inequality. In view of (3.3) it suffices to show

$$\text{(3.9)} \qquad \text{a.s.} \quad \limsup_n \frac{\sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_i}\right) I_{x - h_i S_k}(X_i)}{1 + \sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)} \leq c^* m(x) \bmod \mu$$

for some $c^* \in \mathbb{R}_+$ independent of $P_{(X,Y)}$ with $Y \geq 0$, $EY < \infty$. Without loss of generality in the following $K \leq 1$ may be assumed.

In the second step, according to Lemma 3.2 with $r_n := \int K(\frac{x-t}{h_n}) I_{t + h_n S_k}(x)\mu(dx)$ we choose for $t \in \mathbb{R}^d$ indices $p_i = p(t,k,i) \uparrow \infty$ $(i \to \infty)$ such that (3.5), (3.6), (3.7) hold for all $i \in \mathbb{N}$. For $p(t,k,\cdot)$ we define an inverse function $q(t,k,\cdot)$ on $\mathbb{N}$ by $q(t,k,n) := \max\{i \in \mathbb{N}; p(t,k,i) \leq n\}$. Further we define the truncated random variables $Z_i := Y_i I_{[Y_i \leq q(X_i,k,i)]}$, $i \in \mathbb{N}$. It will be shown

$$\text{(3.10)} \quad \text{a.s.} \quad \frac{\sum_{i=1}^n \left[Z_i K\left(\frac{x - X_i}{h_i}\right) I_{x - h_i S_k}(X_i) - EZ_i K\left(\frac{x - X_i}{h_i}\right) I_{x - h_i S_k}(X_i)\right]}{1 + \sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)} \to 0 \bmod \mu.$$

Because of (3.2), according to Section 16.3 in Loève (1977) it suffices to show

$$\sum_n \frac{EZ_n^2 K\left(\frac{x - X_n}{h_n}\right)^2 I_{S_k}\left(\frac{x - X_n}{h_n}\right)}{\left(1 + \sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)\right)^2} < \infty \bmod \mu.$$

But this follows from

$$\int \sum_{n=1}^\infty \frac{EZ_n^2 K\left(\frac{x - X_n}{h_n}\right)^2 I_{S_k}\left(\frac{x - X_n}{h_n}\right)}{\left(1 + \sum_{j=1}^n \int K\left(\frac{s-x}{h_j}\right)\mu(ds)\right)^2}\mu(dx)$$

$$= \sum_{n=1}^\infty \int \left[\int \frac{E(Z_n^2 \mid X_n = t) K\left(\frac{x-t}{h_n}\right)^2 I_{t + h_n S_k}(x)}{\left(1 + \sum_{j=1}^n \int K\left(\frac{s-x}{h_j}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt)$$

$$\le \frac{1}{c^2}\sum_{n=1}^{\infty}\int\left[\int \frac{E(Z_n^2 \mid X_n = t)K\left(\frac{x-t}{h_n}\right) I_{t+h_n S_k}(x)}{\left(1+\sum_{j=1}^n \int K\left(\frac{s-t}{h_j}\right) I_{S_k}\left(\frac{s-t}{h_j}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt)$$

$$= \frac{1}{c^2}\sum_{n=1}^{\infty}\int\left[\int \sum_{i=1}^{q(t,k,n)}\left(\int_{(i-1,i]} v^2 P_{Y\mid X=t}(dv)\right)\right.$$

$$\left.\cdot \frac{K\left(\frac{x-t}{h_n}\right) I_{t+h_n S_k}(x)}{\left(1+\sum_{j=1}^n \int K\left(\frac{s-t}{h_j}\right) I_{S_k}\left(\frac{s-t}{h_j}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt)$$

$$= \frac{1}{c^2}\int \sum_{i=1}^{\infty}\left(\int_{(i-1,i]} v^2 P_{Y\mid X=t}(dv)\right)$$

$$\cdot \sum_{n=p(t,k,i)}^{\infty} \frac{\int K\left(\frac{x-t}{h_n}\right) I_{t+h_n S_k}(x)\mu(dx)}{\left(1+\sum_{j=1}^n \int K\left(\frac{s-t}{h_j}\right) I_{t+h_j S_k}(s)\mu(ds)\right)^2}\mu(dt)$$

$$\le \frac{1}{c^2}\int E(Y \mid X = t)\mu(dt)$$

$$= \frac{1}{c^2} EY < \infty,$$

where we obtain the first inequality from (3.8) and the second inequality from (3.6).

In the third step we notice

$$\limsup_n \frac{\sum_{i=1}^n EZ_i K\left(\frac{x-X_i}{h_i}\right) I_{x-h_i S_k}(X_i)}{1+\sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)} \le \lim_n \frac{\sum_{i=1}^n \int m(t) K\left(\frac{x-t}{h_i}\right)\mu(dt)}{1+\sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)}$$
$$= m(x) \bmod \mu$$

because of (3.1), (3.2) and the Toeplitz theorem. This together with (3.10) yields

$$\text{(3.11)} \qquad \text{a.s.} \qquad \limsup_n \frac{\sum_{i=1}^n Z_i K\left(\frac{x-X_i}{h_i}\right) I_{x-h_i S_k}(X_i)}{1+\sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)} \le m(x) \bmod \mu.$$

In the fourth step we use the special assumption $\alpha I_S \le K \le \beta I_S$ with $0 < \alpha < \beta < \infty$ and show

$$\text{(3.12)} \qquad \sum_n P\left[Z_n I_{S\cap S_k}\left(\frac{x-X_n}{h_n}\right) \ne Y_n I_{S\cap S_k}\left(\frac{x-X_n}{h_n}\right)\right] < \infty \bmod \mu.$$

But this follows from

$$\int \sum_{n=1}^{\infty} P[Y_n > q(X_n, k, n), X_n \in x - h_n(S \cap S_k)]\mu(dx)$$
$$= \int \sum_{n=1}^{\infty} \left(\int P\left[Y > q(t,k,n) \mid X = t\right] I_{x-h_n(S\cap S_k)}(t)\mu(dt) \right) \mu(dx)$$
$$= \sum_{n=1}^{\infty} \int P[Y > q(t,k,n) \mid X = t]\mu(t + h_n(S \cap S_k))\mu(dt)$$
$$\leq \int \sum_{i=1}^{\infty} P[Y \in (i, i+1] \mid X = t] \sum_{n=1}^{p(t,k,i+1)} \mu(t + h_n(S \cap S_k))\mu(dt)$$
$$\leq 3 \int E(Y \mid X = t)\mu(dt)$$
$$= 3EY < \infty$$

by use of (3.5). Relation (3.12) yields that a.s. for μ-almost all x from some random index N_x on

$$Z_n I_{S\cap S_k}\left(\frac{x - X_n}{h_n}\right) = Y_n I_{S\cap S_k}\left(\frac{x - X_n}{h_n}\right).$$

This together with (3.11) and (3.2) yields (3.9). Thus the assertion in a) is proved.

In the fifth step, in the context with general K we use the assumption $EY \ln^+ Y < \infty$. We show

$$\text{(3.13)} \qquad \text{a.s.} \quad \frac{\sum_{i=1}^{n}(Y_i - Z_i)K\left(\frac{x - X_i}{h_i}\right) I_{x-h_iS_k}(X_i)}{1 + \sum_{i=1}^{n} \int K\left(\frac{x-t}{h_i}\right)\mu(dt)} \to 0 \text{ mod } \mu.$$

By (3.2) and the Toeplitz theorem, this follows from

$$\int \sum_{n=1}^{\infty} \frac{EY_n I_{[Y_n > q(X_n,k,n)]} K\left(\frac{x - X_n}{h_n}\right) I_{X_n + h_n S_k}(x)}{1 + \sum_{i=1}^{n} \int K\left(\frac{t-x}{h_i}\right)\mu(dt)} \mu(dx)$$
$$= \int \sum_{n=1}^{\infty} \frac{\int E(Y I_{[Y > q(t,k,n)]} \mid X = t) K\left(\frac{x-t}{h_n}\right) I_{t+h_nS_k}(x)\mu(dt)}{1 + \sum_{i=1}^{n} \int K\left(\frac{t-x}{h_i}\right)\mu(dt)} \mu(dx)$$
$$= \sum_{n=1}^{\infty} \int \frac{\sum_{i=q(t,k,n)}^{\infty} \int_{(i,i+1]} v P_{Y|X=t}(dv) K\left(\frac{x-t}{h_n}\right) I_{t+h_nS_k}(x)\mu(dt)}{1 + \sum_{i=1}^{n} \int K\left(\frac{s-x}{h_i}\right)\mu(ds)} \mu(dx)$$
$$\leq \frac{1}{c} \int \sum_{i=1}^{\infty} \left[\int_{(i,i+1]} v P_{Y|X=t}(dv) \right]$$

$$\cdot \sum_{n=1}^{p(t,k,i+1)} \frac{\int K\left(\frac{x-t}{h_n}\right) I_{t+h_n S_k}(x)\mu(dx)}{1+\sum_{i=1}^n \int K\left(\frac{s-t}{h_i}\right) I_{t+h_i S_k}(s)\mu(ds)} \mu(dt)$$
$$\leq \frac{1}{c}\int \sum_{i=1}^{\infty} \int_{(i,i+1]} v P_{Y|X=t}(dv)\mu(dt)\ln(i+3)$$
$$\leq \frac{1}{c}(3+EY\ln^+ Y) < \infty,$$

where we obtain the first inequality from (3.8) and the second inequality from (3.7). (3.13) together with (3.11) yields (3.9). Thus the assertion in b) is proved. □

PROOF OF THEOREM 2.2. We use Lemma 3.1 with $K_n(x,t) = I_{A_n(x)}(t)$.

Equations (3.1) and (3.2) follow from (13) and (14) in the proof of Theorem 2 in Györfi *et al.* (1998). A simple proof of (3.1) is given in Algoet and Györfi ((1999), p. 136): $(f_n, \mathcal{P}_n)_{n\in\mathbb{N}}$ with $f_n(x) := E(Y \mid X \in A_n(x))$, $x \in \mathbb{R}^d$, is a martingale on $(\mathbb{R}^d, \mathcal{B}_d, \mu)$ since $(\mathcal{P}_n)$ is nested. By the martingale convergence theorem $(f_n(x))$ converges modμ, where by diam $A_n(x) \to 0$, $x \in \mathbb{R}^d$, the limit is $m(x)$ modμ. This argument and the proof of Lemma 2.2 in Devroye (1981) also yield (3.2): for a fixed open sphere S_R in $\mathbb{R}^d$ around 0 with radius R, $(g_n, \mathcal{P}_n)_{n\in\mathbb{N}}$ with $g_n(x) := \lambda(A_n(x)\cap S_R)/\mu(A_n(x)\cap S_R)$, $x \in S_R$, is a martingale on $(S_R, S_R \cap \mathcal{B}_d, \mu_{|S_R\cap\mathcal{B}_d})$, where on S_R $(g_n(x))$ converges mod μ (for application of the martingale convergence theorem notice $\lambda(S_R) < \infty$); thus, by diam $A_n(x) \to 0$, $x \in \mathbb{R}^d$, it holds on S_R, even on $\mathbb{R}^d$, $\lim \mu(A_n(x))/\lambda(A_n(x)) > 0$ mod μ, which together with $\sum \lambda(A_n(x)) = \infty$, $x \in \mathbb{R}^d$, leads to (3.2).

It remains to verify (3.3) for $P_{(X,Y)}$ with $Y \geq 0$, $EY < \infty$. According to Lemma 3.2 with $r_n = \mu(A_n(t))$ we choose for $t \in \mathbb{R}^d$ indices $p_i = p(t,i) \uparrow \infty$ $(i \to \infty)$ such that (3.5), (3.6) hold for all $i \in \mathbb{N}$. As in the proof of Theorem 1, we define for $p(t,\cdot)$ an inverse function $q(t,\cdot)$ on $\mathbb{N}$ by $q(t,n) := \max\{i \in \mathbb{N}; p(t,i) \leq n\}$, further the truncated random variables $Z_i := Y_i I_{[Y_i \leq q(X_i,i)]}$, $i \in \mathbb{N}$. For the nested sequence of partitions, as in the proof of Theorem 2 in Györfi *et al.* (1998) we notice that $x \in A_n(t)$ and $j \leq n$ imply $A_j(x) = A_j(t)$. Now as in the proof of Theorem 3.1 we obtain

$$\text{a.s.} \quad \frac{\sum_{i=1}^n [Z_i I_{A_i(x)}(X_i) - EZ_i I_{A_i(x)}(X_i)]}{1+\sum_{i=1}^n \mu(A_i(x))} \to 0 \bmod \mu.$$

Further

$$\limsup_n \frac{\sum_{i=1}^n EZ_i I_{A_i(x)}(X_i)}{1+\sum_{i=1}^n \mu(A_i(x))} \leq \lim_n \frac{\sum_{i=1}^n \int m(t) I_{A_i(x)}(t)\mu(dt)}{1+\sum_{i=1}^n \mu(A_i(x))} = m(x) \bmod \mu$$

because of (3.1), (3.2) and the Toeplitz theorem. Thus

$$\text{(3.14)} \qquad \text{a.s.} \quad \limsup_n \frac{\sum_{i=1}^n Z_i I_{A_i(x)}(X_i)}{1+\sum_{i=1}^n \mu(A_i(x))} \leq m(x) \bmod \mu.$$

As in the proof of Theorem 2.1, we obtain

$$\text{(3.15)} \qquad \sum_n P[Z_n I_{A_n(x)}(X_n) \neq Y_n I_{A_n(x)}(X_n)] < \infty \bmod \mu.$$

Because of (3.2),

$$1 + \sum_{i=1}^{n} \mu(A_i(x)) \to \infty \text{ mod } \mu. \tag{3.16}$$

(3.14), (3.15) and (3.16) yield (3.3). Now the assertion follows by Lemma 3.1. □

PROOF OF THE PROPOSITION. Without loss of generality $Y_n \geq 0$ may be assumed. We use the notations

$$\begin{aligned} A_n(x) &:= a_n \bar{K}_n(x, X_n), \\ B_n(x) &:= [(1 - A_2(x)) \dots (1 - A_n(x))]^{-1}, \\ G_n(x) &:= A_n(x) B_n(x) \qquad (n = 2, 3, \dots), \\ B_1(x) &:= 1, \qquad G_1(x) := 1. \end{aligned}$$

Representations of the following kind are well known (compare Ljung *et al.* (1992), part I, Lemma 1.1):

$$B_n(x) = \sum_{i=1}^{n} G_i(x) \ (n \in \mathbb{N}), \tag{3.17}$$

$$m_n(x) = B_n(x)^{-1} \sum_{i=1}^{n} G_i(x) Y_i \ (n \in \mathbb{N}). \tag{3.18}$$

We notice a.s. $\sum A_n(x) = \infty$ mod μ by a.s. convergence of $\sum(A_n(x) - EA_n(x))$ (because of (2.4) and thus $\sum EA_n(x)^2 < \infty$) and $\sum EA_n(x) = \infty$ mod μ (because of (2.3) and (2.2)), thus

$$\text{a.s. } B_n(x) \uparrow \infty \text{ mod } \mu. \tag{3.19}$$

Let $\tilde{Y}_n := Y_n I_{[Y_n \leq n]} (n \in \mathbb{N})$. As is well known from the proof of Kolmogorov's strong law of large numbers for independent identically distributed integrable random variables,

$$\sum \frac{1}{n^2} E\tilde{Y}_n^2 < \infty \tag{3.20}$$

and

$$\sum P[Y_n \neq \tilde{Y}_n] < \infty. \tag{3.21}$$

By (3.18) we can use the representation $m_n(x) = m_n^{(1)}(x) + m_n^{(2)}(x) + m_n^{(3)}(x)$, $n \in \mathbb{N}$, with

$$\begin{aligned} m_n^{(1)}(x) &= \frac{1}{B_n(x)} \sum_{i=1}^{n} G_i(x) \left(\tilde{Y}_i - \frac{E\tilde{Y}_i \bar{K}_i(x, X_i)}{E\bar{K}_i(x, X_i)} \right), \\ m_n^{(2)}(x) &= \frac{1}{B_n(x)} \sum_{i=1}^{n} G_i(x) \frac{E\tilde{Y}_i \bar{K}_i(x, X_i)}{E\bar{K}_i(x, X_i)}, \\ m_n^{(3)}(x) &= \frac{1}{B_n(x)} \sum_{i=1}^{n} G_i(x)(Y_i - \tilde{Y}_i). \end{aligned}$$

In the first step we show

$$\text{(3.22)} \qquad \text{a.s.} \quad m_n^1(x) \to 0 \bmod \mu.$$

By (3.19) and the Kronecker lemma it suffices to show a.s. convergence of

$$\sum A_n(x)\left(\tilde{Y}_n - \frac{E\tilde{Y}_n\bar{K}_n(x,X_n)}{E\bar{K}_n(x,X_n)}\right).$$

But this follows from

$$\sum E(A_n(x)\tilde{Y}_n)^2 < \infty,$$

which holds because of (2.4) and (3.20).

In the second step we show

$$\text{(3.23)} \qquad \text{a.s.} \quad m_n^{(2)}(x) \to m(x) \bmod \mu.$$

Because of (3.17), (3.19) and the Toeplitz theorem it suffices to show

$$\text{(3.24)} \qquad \frac{E\tilde{Y}_n\bar{K}_n(x,X_n)}{E\bar{K}_n(x,X_n)} \to m(x) \bmod \mu.$$

By (2.1) we have

$$\limsup_n \frac{E\tilde{Y}_n\bar{K}_n(x,X_n)}{E\bar{K}_n(x,X_n)} \le \lim_n \frac{EY\bar{K}_n(x,X)}{E\bar{K}_n(x,X)} = m(x) \bmod \mu,$$

on the other side for each $c \in \mathbb{N}$

$$\liminf_n \frac{E\tilde{Y}_n\bar{K}_n(x,X_n)}{E\bar{K}_n(x,X_n)} \ge \lim_n \frac{EYI_{[Y\le c]}\bar{K}_n(x,X)}{E\bar{K}_n(x,X)} = E(YI_{[Y\le c]} \mid X = x) \bmod \mu.$$

These relations with $c \to \infty$ yield (3.24).

In the third step we obtain

$$\text{(3.25)} \qquad \text{a.s.} \quad m_n^{(3)}(x) \to 0 \bmod \mu$$

by (3.19) and (3.21). Now (3.22) , (3.23), (3.25) yield the assertion. □

PROOF OF THEOREM 2.3. Setting

$$\bar{K}_n(x,z) = \frac{1}{h_n^d}K\left(\frac{x-z}{h_n}\right), \qquad (x,z) \in \mathbb{R}^d \times \mathbb{R}^d, \quad n \in \{2,3,\ldots\},$$

we verify the conditions of the proposition. (2.1) and (2.2) follow from Lemma 1 in Greblicki *et al.* (1984) and from the proof of Lemma 2.2 of Devroye (1981), respectively. □

PROOF OF THEOREM 2.4. Setting

$$\bar{K}_n(x,z) = \frac{1}{h_n^d}I_{A_n(x)}(z), \qquad (x,z) \in \mathbb{R}^d \times \mathbb{R}^d, \quad n \in \{2,3,\ldots\},$$

we verify the conditions of the proposition. As in the beginning of the proof of Theorem 2.2 we obtain (2.1), also $\lim \mu(A_n(x))/\lambda(A_n(x)) > 0 \bmod \mu$ and thus (2.2). □

PROOF OF REMARK 2.4. The proof differs from that of Theorems 2.2 and 2.4 only at the beginning, namely in the proof of relation

$$\int_{A_n(x)} f(z)\mu(dz)/\mu(A_n(x)) \to f(x) \bmod \mu, \tag{3.26}$$

i.e. (3.1) with $K_n(x,t) = I_{A_n(x)}(t)$, for all μ-integrable functions f on $\mathbb{R}$, and of relation

$$\liminf_n \mu(A_n(x))/\lambda(A_n(x)) > 0 \bmod \mu. \tag{3.27}$$

Let ν be a finite measure on the Borel σ-algebra $\mathcal{B}$ in $\mathbb{R}$. For each $\alpha > 0$ we set

$$M := M_\alpha := \left\{ x \in \mathbb{R}; \sup_n \nu(A_n(x))/\mu(A_n(x)) > \alpha \right\}$$

and shall show

$$\mu(M) \leq \frac{2}{\alpha}\nu(\mathbb{R}) \tag{3.28}$$

by a modification of a covering argument of de Guzmán (1970) and of the proof of Lemma (10.47a) in Wheeden and Zygmund (1977). Let G be an arbitrary bounded subset of $\mathbb{R}$ and set

$$D_N := \left\{ x \in G \cap M : \underset{n\in\{1,\dots,N\}}{\exists} \nu(A_n(x))/\mu(A_n(x)) > \alpha \right\}, \qquad N \in \mathbb{N}.$$

Then $D_N \uparrow G \cap M$. Let $N \in \mathbb{N}$ be arbitrary. For each $n \in \mathbb{N}$ the set D_N is covered by a finite subfamily $\mathcal{P}'_n$ of $\mathcal{P}_n$. For each $x \in D_N$ choose $n_x \in \{1,\dots,N\}$ with $\nu(A_{n_x}(x))/\mu(A_{n_x}(x)) > \alpha$. The intervals $A(x) := A_{n_x}(x)$, $x \in D_N$, form a (finite) subfamily $\mathcal{Q}_N$ of $\cup_{n=1}^N \mathcal{P}'_n$ and cover D_N. We use the following selection procedure. First choose the interval A^1 from $\mathcal{Q}_N$ with largest extension to the left, if not unique among the possible intervals that with largest extension to the right. Let $A^1,\dots,A^k$ already be chosen from $\mathcal{Q}_N$ and let $\mathcal{Q}_{N,k}$ be the subfamily of $\mathcal{Q}_N \backslash \{A^1,\dots,A^k\}$ consisting of the intervals which have a larger extension to the right than A^k. If $\mathcal{Q}_{N,k}$ is void, then stop the selection procedure. If $\mathcal{Q}_{N,k}$ is non-void and if its subfamily $\mathcal{Q}'_{N,k}$ of intervals A' for which $A' \cup A^k$ is also an interval, is non-void, then choose an interval A^{k+1} from $\mathcal{Q}'_{N,k}$ with largest extension to the right. If $\mathcal{Q}_{N,k}$ is non-void and $\mathcal{Q}'_{N,k}$ is void, then choose an interval A^{k+1} from $\mathcal{Q}_{N,k}$ according to the rule for choice of A^1 from $\mathcal{Q}_N$. The procedure stops after a finite number l of steps and yields intervals $A^1,\dots,A^l \in \mathcal{Q}_N$ with $\nu(A^j)/\mu(A^j) > \alpha (j = 1,\dots,l)$ such that $A^1,\dots,A^l$ cover D_N and that each $x \in \mathbb{R}$ is covered at most two times by these intervals. Thus

$$\mu(D_N) \leq \sum_{j=1}^{l} \mu(A^j) < \frac{1}{\alpha}\sum_{j=1}^{l} \nu(A^j) \leq \frac{2}{\alpha}\nu\left(\bigcup_{j=1}^{l} A^j\right) \leq \frac{2}{\alpha}\nu(\mathbb{R})$$

and, by $N \to \infty$, $\mu(G \cap M) \leq \frac{2}{\alpha}\nu(\mathbb{R})$. Letting $G \uparrow \mathbb{R}$ we obtain (3.28). Now from (3.28) with $\nu(B) = \int_B f(x)\mu(dx)$, $B \in \mathcal{B}$, and diam $A_n(z) \to 0$, $z \in \mathbb{R}$, we obtain (3.26) as in the proof of Theorem (10.49) in Wheeden and Zygmund (1977). As to the proof of (3.27), let S_R be the open R-neighborhood of 0 and define the finite measure λ' by $\lambda'(B) = \lambda(B \cap S_R)$, $B \in \mathcal{B}$. Because of diam $A_n(z) \to 0$, $z \in \mathbb{R}$, we obtain

$$\begin{aligned}
\mu &\left(\left\{x \in S_R; \limsup_n \lambda(A_n(x))/\mu(A_n(x)) = \infty\right\}\right) \\
&= \mu\left(\left\{x \in S_R; \limsup_n \lambda'(A_n(x))/\mu(A_n(x)) = \infty\right\}\right) \\
&\le \mu\left(\left\{x \in \mathbb{R}; \sup_n \lambda'(A_n(x))/\mu(A_n(x)) = \infty\right\}\right) \\
&= \lim_{\alpha\to\infty} \mu\left(\left\{x \in \mathbb{R}; \sup_n \lambda'(A_n(x))/\mu(A_n(x)) > \alpha\right\}\right) \\
&= 0
\end{aligned}$$

by (3.28) with $\nu = \lambda'$, and thus, by $R \to \infty$, relation (3.27). □

Acknowledgements

The author thanks P. Algoet and L. Györfi for valuable discussions and the referees whose suggestions improved the paper.

References

Ahmad, I. A. and Lin, P.-E. (1976). Nonparametric sequential estimation of a multiple regression function, *Bulletin of Mathematical Statistics*, **17**, 63–75.

Algoet, P. and Györfi, L. (1999). Strong universal pointwise consistency of some regression function estimates, *J. Multivariate Anal.*, **71**, 125–144.

de Guzmán, M. (1970). A covering lemma with applications to differentiability of measures and singular integral operators, *Studia Math.*, **34**, 299–317.

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

Devroye, L. and Györfi, L. (1983). Distribution-free exponential bound on the L_1 error of partitioning estimates of a regression function, *Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics* (eds. F. Konecny, J. Mogyoródi, and W. Wertz), 67–76, Akadémiai Kiadó, Budapest.

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

Devroye, L. and Wagner, T. J. (1980*a*). Distribution-free consistency results in nonparametric discrimination and regression function estimation, *Ann. Statist.*, **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, *Z. Wahrsch. Verw. Gebiete*, **51**, 15–25.

Greblicki, W. (1974). Asymptotically optimal probabilistic algorithms for pattern recognition and identification, *Prace Nauk. Inst. Cybernet. Tech. Politech. Wroclaw.*, No. 18., Ser. Monogr., No. 3.

Greblicki, W. and Pawlak, M. (1985). Fourier and Hermite series estimates of regession functions, *Ann. Inst. Statist. Math.*, **37**, 443–454.

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

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

Györfi, L. (1991). Universal consistencies of a regression estimate for unbounded regression functions, *Nonparametric Functional Estimation* (ed. G. Roussas), *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, 329–338, Kluwer, Dordrecht.

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

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

Knopp, K. (1956). *Infinite Sequences and Series*, Dover, New York.

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

Krzyżak, A. and Pawlak, M. (1984). Almost everywhere convergence of a recursive regression function estimate and classification, *IEEE Trans. Inform. Theory*, **IT-30**, 91–93.

Ljung, L., Pflug, G. and Walk, H. (1992). *Stochastic Approximation and Optimization of Random Systems*, Birkhäuser, Basel.

Loève, M. (1977). *Probability I*, 4th ed., Springer, Berlin, Heidelberg, New York.

Nadaraya, E. A. (1964). On estimating regression, *Theory Probab. Appl.*, **9**, 141–142.

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

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

Stone, C. J. (1977). Consistent nonparametric regression, *Ann. Statist.*, **5**, 595–645.

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

Walk, H. (2000). Almost sure convergence properties of Nadaraya-Watson regression estimates, *Modeling Uncertainty: An Examination of its Theory, Methods and Applications* (S. Yakowitz memorial volume) (eds. M. Dror, P. L'Ecuyer and F. Szidarovszky), Kluwer, Dordrecht (to appear).

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

Wheeden, R. L. and Zygmund, A. (1977). *Measure and Integral*, Marcel Dekker, New York.

Wolverton, C. T. and Wagner, T. J. (1969). Recursive estimates of probability densities, *IEEE Transactions on Systems, Man, and Cybernetics*, **5**, 307.

Yamato, H. (1971). Sequential estimation of a continuous probability density function and mode, *Bulletin of Mathematical Statistics*, **14**, 1–12.