Universität Stuttgart

Fachbereich Mathematik

Strong Laws of Large Numbers and Nonparametric Estimation

Harro Walk

Preprint 2009/003

Universität Stuttgart
Fachbereich Mathematik

Strong Laws of Large Numbers and Nonparametric Estimation

Harro Walk

Preprint 2009/003

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik/uni-stuttgart.de/preprints

ISSN **1613-8309**

Strong Laws of Large Numbers and Nonparametric Estimation

Harro Walk

Revidierte Fassung vom 29.04.2010

Abstract Elementary approaches to classic strong laws of large numbers use a monotonicity argument or a Tauberian argument of summability theory. Together with results on variance of sums of dependent random variables they allow to establish various strong laws of large numbers in case of dependence, especially under mixing conditions. Strong consistency of nonparametric regression estimates of local averaging type (kernel and nearest neighbor estimates), pointwise as well as in L_2, can be considered as a generalization of strong laws of large numbers. Both approaches can be used to establish strong universal consistency in the case of independence and, mostly by sharpened integrability assumptions, consistency under ρ-mixing or α-mixing. In a similar way Rosenblatt-Parzen kernel density estimates are treated.

1 Introduction

The classic strong law of large numbers of Kolmogorov deals with independent identically distributed integrable real random variables. An elementary approach has been given by Etemadi (1981). He included the arithmetic means of nonnegative (without loss of generality) random variables truncated at the number equal to the index, between fractions with the first $\lceil a^{N+1} \rceil$ summands in the numerator and the denominator a^N and fractions with the first $\lfloor a^N \rfloor$ summands in the numerator and the denominator a^{N+1} ($a > 1$, rational), investigated the almost sure (a.s.) convergence behavior of the majorant sequence and the minorant sequence by use of Chebyshev's inequality and let then go $a \downarrow 1$. This method was refined by Etemadi (1983) himself, Csörgő, Tandori and Totik (1983) and Chandra and Goswami (1992, 1993) and extended to the investigation of nonparametric regression and density estimates under mixing conditions by Irle (1997) and to the proof of strong universal pointwise consistency of nearest neighbor regression estimates under independence by Walk (2008a).

Another approach to strong laws of large numbers was proposed by Walk (2005b). Classic elementary Tauberian theorems (Lemma 1a,b) in summmability theory allow to conclude convergence of a sequence (s_n) of real numbers from convergence of their arithmetic means (C_1 summability, Cesàro summability of (s_n)) together with a so-called Tauberian condition on variation of the original sequence (s_n). If (s_n) itself is a sequence of basic arithmetic means $(a_1 + \cdots + a_n)/n$, as the realization in the strong law of large numbers, then the Tauberian condition simply means that (a_n) is bounded from below. In this context the other assumption (C_1-summability of the sequence of basic arithmetic means) is usually replaced by the more practicable, but equivalent, C_2-summability of (a_n), see Lemma 1a. For the sequence of nonnegative truncated random variables centered at their expectations which are bounded by the finite expectation in Kolmogorov's strong law of large numbers, the simple Tauberian condition is obviously fulfilled. To show almost sure (a.s.) C_2-summability of the sequence to 0, it then suffices to show a.s. convergence of a series of nonnegative random variables by taking expectations, see Theorem 1a. The summmability theory approach has been extended by Walk (2005a, 2008b) to establish strong universal L_2-consistency

Harro Walk, Universität Stuttgart, Department of Mathematics, Pfaffenwaldring 57, D-70569 Stuttgart, e-mail: harro.walk@mathematik.uni-stuttgart.de

of Nadaraya-Watson type kernel regression estimates (under independence) and strong consistency under mixing conditions and sharpened moment conditions. − Both described approaches have different areas of application and will be used in this paper.

In Section 2 strong laws of large numbers under conditions on the covariance and more generally under conditions on the variance of sums of random variables (Theorem 1) and under mixing conditions (Theorem 2) are stated. For the two latter situations proofs via the summability theory approach are given. We shall deal with ρ-mixing and α-mixing conditions. Theorem 2a specialized to the case of independence states Kolmogorov's strong law of large numbers and is a consequence of Theorem 1a, which itself is an immediate consequence of the Tauberian Lemma 1a.

Section 3 deals with strong pointwise consistency of Nadaraya-Watson kernel regression estimates under ρ-mixing and α-mixing (Theorem 3), Devroye-Wagner semirecursive kernel regression estimates under ρ-mixing (Theorem 4) and k_n-nearest neighbor regression estimates under independence (Theorem 5). In the proof of Theorem 3 truncation of the response variables is justified by a monotonicity argument of Etemadi type, asymptotic unbiasedness is established by a generalized Lebesgue density theorem of Greblicki, Krzyżak and Pawlak (1984), and a.s. convergence after truncation and centering at expectations is shown by exponential inequalities of Peligrad (1992) and Rhomari (2002). Theorem 4 is a result on strong universal pointwise consistency, i.e., strong pointwise consistency for each distribution of (X,Y) with $\mathbf{E}|Y| < \infty$ (X d-dimensional prediction random vector, Y real response random variable); it is an extension from independence (Walk (2001)) to ρ-mixing due to Frey (2007) by use of the Tauberian Lemma 2 on weighted means. Theorem 5 is a strong universal consistency result of Walk (2008a). Its proof uses Etemadi's monotonicity argument and will be omitted. Irle (1997) uses mixing and boundedness assumptions (Remark 5). Section 4 first points out strong universal L_2-consistency (strong L_2-consistency for all distributions of (X,Y) with $\mathbf{E}\{|Y|^2\} < \infty$ under independence) of k_n-nearest neighbor, semirecursive Devroye-Wagner kernel and Nadaraya-Watson type kernel estimates (Devroye et al. (1994), Györfi, Kohler and Walk (1998) and Walk (2005a), respectively), see Theorem 6 (without proof). Under ρ- and α-mixing and sharpened moment conditions, Theorem 7 (Walk 2008b) states strong L_2- consistency of Nadaraya-Watson regression estimates. Its proof uses the summability theory approach and will be omitted.

The final Section 5 deals with Rosenblatt-Parzen kernel density estimates under ρ- and α-mixing. L_1-consistency (Theorem 8) is proven by use of a monotonicity argument of Etemadi type.

2 Strong laws of large numbers

The following lemma states elementary classical Tauberian theorems of Landau (1910) and Schmidt (1925). They allow to conclude Cesàro summability (C_1-summability, i.e., convergence of arithmetic means) from C_2-summability or to conclude convergence from C_1-summability, in each case under an additional assumption (so-called Tauberian condition). A corresponding result of Szász (1929) and Karamata (1938) concerns weighted means (Lemma 2). References for these and related results are Hardy (1949), pp. 121, 124-126, 145, Zeller and Beekmann (1970), pp. 101, 103, 111-113, 117, Korevaar (2004), pp. 12-16, 58, compare also Walk (2005b, 2007).

Lemma 1. *a) Let the sequence $(a_n)_{n\in\mathbb{N}}$ of real numbers satisfy*

$$\frac{1}{\binom{n+1}{2}} \sum_{j=1}^{n} \sum_{i=1}^{j} a_i \to 0, \tag{1}$$

i.e., C_2-summability of $(a_n)_{n\in\mathbb{N}}$ to 0, or sharper

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \left(\sum_{i=1}^{n} a_i \right)^2 < \infty, \tag{2}$$

and the Tauberian condition

$$\inf_n a_n > -\infty. \tag{3}$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} a_i \to 0. \tag{4}$$

b) Let the sequence $(s_n)_{n\in\mathbb{N}}$ *of real numbers satisfy*

$$\frac{1}{n}\sum_{k=1}^{n} s_k \to 0 \tag{5}$$

and the Tanberian condition

$$\liminf\ (s_N - s_M) \geq 0 \quad for\ M \to \infty,\ M < N,\ N/M \to 1, \tag{6}$$

i.e.,

$$\liminf\ (s_{N_n} - s_{M_n}) \geq 0$$

for each sequence $((M_n, N_n))$ *in* $\mathbb{N}^2$ *with* $M_n \to \infty$, $M_n < N_n$, $N_n/M_n \to 1$ $(n \to \infty)$.
Then

$$s_n \to 0. \tag{7}$$

To make the paper more self-contained we shall give direct proofs of Lemma 1a and Lemma 1b. Remark 1b states (with proof) that Lemma 1b implies Lemma 1a. The notations $\lfloor s \rfloor$ and $\lceil s \rceil$ fot the integer part and the upper integer part of the nonnegative real number s will be used.

Proof (of Lemma 1).

a) (2) implies (1), because

$$\left|\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{j} a_i\right|^2 \leq \frac{1}{n^4} n \sum_{j=1}^{n}\left(\sum_{i=1}^{j} a_i\right)^2 \to 0 \quad (n \to \infty)$$

by the Cauchy-Schwarz inequality, (2) and the Kronecker lemma. (3) means $a_n \geq -c$, $n \in \mathbb{N}$, for some $c \in \mathbb{R}_+$. With

$$t_n := \sum_{i=1}^{n} a_i, \quad w_n := \sum_{j=1}^{n} t_j, \quad n \in \mathbb{N},$$

for $k \in \{1,\ldots,n\}$ one obtains

$$w_{n+k} - w_n = t_n k + \sum_{j=n+1}^{n+k} (t_j - t_n) \geq k t_n - k^2 c,$$

$$w_{n-k} - w_n = t_n(-k) + \sum_{j=n-k+1}^{n} (t_n - t_j) \geq -k t_n - k^2 c$$

(compare Taylor expansion), thus

$$\frac{w_n - w_{n-k}}{nk} - \frac{kc}{n} \leq \frac{t_n}{n} \leq \frac{w_{n+k} - w_n}{nk} + \frac{kc}{n}.$$

(1) implies

$$\sigma_n := \max\{|w_l|; l = 1,\ldots,2n\} = o(n^2),$$
$$k(n) := \min\{1 + \lfloor\sqrt{\sigma_n}\rfloor, n\} = o(n).$$

Therefore

$$\begin{aligned}\frac{|t_n|}{n} &\le \frac{2\sigma_n}{nk(n)} + \frac{k(n)}{n}c \\ &= \frac{k(n)}{n}\frac{2\sigma_n}{k(n)^2} + \frac{k(n)}{n}c \le (2+c)\frac{k(n)}{n} \to 0 \quad (n \to \infty),\end{aligned}$$

i.e. (4).

b) With

$$z_n := \sum_{k=1}^{n} s_k, \quad n \in \mathbb{N},$$

for $k \in \{1,\dots,\lceil \frac{n}{2} \rceil\}$ we obtain as before

$$z_{n+k} - z_n = s_n k + \sum_{j=n+1}^{n+k} (s_j - s_n),$$

$$z_{n-k} - z_n = s_n(-k) + \sum_{j=n-k+1}^{n} (s_n - s_j),$$

thus for $n \ge 2$

$$\begin{aligned}&-\frac{2}{k/n} \sup_{j\in\{\lfloor \frac{n}{2}\rfloor, \lfloor \frac{n}{2}\rfloor+1,\dots\}} \frac{|z_j|}{j} + (s_n - s_{j(n,k)}) \\ &\le \frac{n}{k}\frac{z_n}{n} - \frac{n-k}{k}\frac{z_{n-k}}{n-k} + \min_{j\in\{n-k+1,\dots,n\}} (s_n - s_j) \\ &\le s_n \\ &\le \frac{n+k}{k}\frac{z_{n+k}}{n+k} - \frac{n}{k}\frac{z_n}{n} - \min_{j\in\{n+1,\dots,n+k\}} (s_j - s_n) \\ &\le 2\frac{1+k/n}{k/n} \sup_{j\in\{n,n+1,\dots\}} \frac{|z_j|}{j} - (s_{j^*(n,k)} - s_n)\end{aligned}$$

with suitable $j(n,k) \in \{n-k+1,\dots,n\}$, $j^*(n,k) \in \{n+1,\dots,n+k\}$. Now choose $k = k(n) \in \{1,\dots,\lceil \frac{n}{2} \rceil\}$ such that $k(n)/n \to 0$ so slowly that, besides $\sup_{j\in\{\lfloor \frac{n}{2}\rfloor, \lfloor \frac{n}{2}\rfloor+1,\dots\}} \frac{|z_j|}{j} \to 0$ $(n \to \infty)$ (by (5)), even

$$\frac{1}{\frac{k(n)}{n}} \sup_{j\in\{\lfloor \frac{n}{2}\rfloor, \lfloor \frac{n}{2}\rfloor+1,\dots\}} \frac{|z_j|}{j} \to 0 \quad (n \to \infty).$$

Therefore and by $k(n)/n \to 0$ (once more) together with (6) we obtain

$$\begin{aligned}0 &\le \liminf(s_n - s_{j(n,k(n))}) \\ &\le \liminf s_n \le \limsup s_n \\ &\le -\liminf(s_{j^*(n,k(n))} - s_n) \le 0,\end{aligned}$$

which yields (7). □

Remark 1.

a) Assumption (6) in Lemma 1b is fulfilled if

$$s_{n+1} - s_n \ge u_n + v_n + w_n$$

with $u_n = O\left(\frac{1}{n}\right)$, convergence of $\left(\frac{1}{n}\sum_{i=1}^{n} i v_i\right)$, $\sum_{n=1}^{\infty} n w_n^2 < \infty$.
For

$$\Big|\sum_{n=M+1}^{N} u_n\Big| \le \text{const} \sum_{n=M+1}^{N} \frac{1}{n} \to 0,$$

$$
\begin{aligned}
\Big|\sum_{n=M+1}^{N} v_n\Big| &= \Big|\sum_{n=M+1}^{N} \frac{1}{n}(nv_n)\Big| \\
&= \Big|\frac{1}{N+1}\sum_{n=1}^{N} nv_n - \frac{1}{M+1}\sum_{n=1}^{M} nv_n + \sum_{n=M+1}^{N}\frac{1}{n(n+1)}\sum_{i=1}^{n} iv_i\Big| \\
&\quad \text{(by partial summation)} \\
&\le o(1) + \sup_{n\in\mathbb{N}}\left(\frac{1}{n+1}\Big|\sum_{i=1}^{n} iv_i\Big|\right)\sum_{n=M+1}^{N}\frac{1}{n} \\
&\quad \to 0,
\end{aligned}
$$

and (by the Cauchy-Schwarz inequality)

$$
\begin{aligned}
\left|\sum_{n=M+1}^{N} w_n\right| &= \left|\sum_{n=M+1}^{N} n^{-\frac{1}{2}}(n^{\frac{1}{2}} w_n)\right| \\
&\le \left(\sum_{n=M+1}^{N}\frac{1}{n}\right)^{\frac{1}{2}}\left(\sum_{n=1}^{\infty} nw_n^2\right)^{\frac{1}{2}} \to 0
\end{aligned}
$$

for $M\to\infty$, $M<N$, $N/M\to 1$.

b) Lemma 1b implies Lemma 1a. For, under the assumptions of Lemma 1a, with $s_n := (a_1+\cdots+a_n)/n$ one has

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{n} s_k \\
&= \frac{1}{n(n+1)}\sum_{k=1}^{n}(a_1+\cdots+a_k) + \frac{1}{n}\sum_{k=1}^{n}\frac{1}{k(k+1)}\sum_{j=1}^{k}(a_1+\cdots+a_j) \\
&\quad \text{(by partial summation)} \\
&\quad \to 0 \quad (n\to\infty)
\end{aligned}
$$

by (1), i.e., (5) is fulfilled. Further, with suitable $c\in\mathbb{R}_+$,

$$
\begin{aligned}
s_{n+1}-s_n &= \frac{a_{n+1}}{n} - \left(\frac{1}{n}-\frac{1}{n+1}\right)(a_1+\cdots+a_{n+1}) \\
&\ge -\frac{c}{n} - \frac{1}{n}s_{n+1} \\
&=: u_n + v_n
\end{aligned}
$$

(by (3)), where $u_n = O\left(\frac{1}{n}\right)$ and

$$
\frac{1}{n}\sum_{i=1}^{n} iv_i = -\frac{1}{n}\sum_{i=1}^{n} s_{i+1} \to 0 \quad (n\to\infty),
$$

by (5). Thus, by a), (6) is fulfilled. Now Lemma 1b yields (7), i.e., (4).

c) Analogously to b) one shows that Lemma 1b implies the variant of Lemma 1a where assumption (3) is replaced by $a_n \ge -c_n$, $n\in\mathbb{N}$, for some sequence (c_n) in $\mathbb{R}_+$ with convergence of $\left(\frac{1}{n}\sum_{i=1}^{n} c_i\right)$.

Part a) of the following Theorem 1 immediately follows from Lemma 1a, compare Walk (2005b), see the proof below. Part b) is due to Chandra and Goswami (1992, 1993) and has been shown by a refinement of Etemadi's (1981, 1983) argument. Part c) contains the classic Rademacher-Menchoff theorem and is obtained according to Serfling (1970b), proof of Theorem 2.1 there; its condition can be slightly weakend (see Walk (2007)). $\mathbf{Cov}_+$ denotes the nonnegative part of $\mathbf{Cov}$, i.e., $\max\{0,\mathbf{Cov}\}$.

Theorem 1. *Let (Y_n) be a sequence of square integrable real random variables. If* *a)* $Y_n \geq 0$, $\sup_n \mathbf{E}Y_n < \infty$ *and*

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \mathbf{Var}\Big\{ \sum_{i=1}^{n} Y_i \Big\} < \infty, \tag{8}$$

or

b) $Y_n \geq 0$, $\sup_n \frac{1}{n} \sum_{k=1}^{n} \mathbf{E}Y_k < \infty$ *and*

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{i=1}^{n} \mathbf{Cov}_+(Y_i, Y_n) < \infty \qquad \left(\Leftrightarrow \sum_{n=1}^{\infty} \frac{1}{n^3} \sum_{j=1}^{n} \sum_{i=1}^{j} \mathbf{Cov}_+(Y_i, Y_j) < \infty \right)$$

or

c) $\sum_{n=1}^{\infty} \frac{(\log(n+1))^2}{n^2} \sum_{i=1}^{n} \mathbf{Cov}_+(Y_i, Y_n) < \infty,$

then

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{E}Y_i) \to 0 \quad a.s. \tag{9}$$

Proof (of Theorem 1a). Obviously $(Y_n - \mathbf{E}Y_n)$ is bounded from below. (8) yields

$$\sum_{n=1}^{\infty} \frac{1}{n^3} \left| \sum_{i=1}^{n} (Y_i - \mathbf{E}Y_i) \right|^2 < \infty \quad a.s.$$

Thus (9) follows by Lemma 1a. □

Remark 2. Analogously one can show that in Theorem 1a the condition $\sup_n \mathbf{E}Y_n < \infty$ may be replaced by convergence of the sequence $\left(\frac{1}{n} \sum_{i=1}^{n} \mathbf{E}Y_i \right)$. Instead of Lemma 1a one uses Remark 1c which is based on Lemma 1b.

Theorem 1a and a corresponding theorem for weighted means based on Lemma 2 below allow to apply results on the variance of sums of dependent random variables (see Theorem 2a and Theorem 4, respectively, with proofs). In the special case of independence, Theorem 2a is Kolmogorov's strong law of large numbers, and its proof by Theorem 1a is elementary.

Remark 3. If the square integrable real random variables Y_n satisfy

$$|\mathbf{Cov}(Y_i, Y_j)| \leq r(|i - j|),$$

then

$$\sum_{n=1}^{\infty} \frac{1}{n} r(n) < \infty$$

or in the case of weak stationarity the weakest possible condition

$$\sum_{n=3}^{\infty} \frac{\log\log n}{n \log n} r(n) < \infty$$

imply

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{E}Y_i) \to 0 \quad a.s.$$

(see Walk (2005b) and Gaposhkin (1977), respectively).

Lemma 2 generalizes Lemma 1a and will be applied in Section 3.

Lemma 2. *Let* $0 < \beta_n \uparrow \infty$ *with* $\beta_{n+1}/\beta_n \to 1$ *and set* $\gamma_n := \beta_n - \beta_{n-1}$ $(n \in \mathbb{N})$ *with* $\beta_0 := 0$. *Let* (a_n) *be a sequence of real numbers bounded from below. If*

$$\frac{1}{\beta_n}\sum_{k=1}^{n}\frac{\gamma_k}{\beta_k}\sum_{j=1}^{k}\gamma_j a_j \to 0$$

or sharper

$$\sum_{n=1}^{\infty}\frac{\gamma_n}{\beta_n^3}\left(\sum_{k=1}^{n}\gamma_k a_k\right)^2 < \infty,$$

then

$$\frac{1}{\beta_n}\sum_{i=1}^{n}\gamma_i a_i \to 0.$$

Also Chandra and Goswami (1992, 1993) gave their result in a more general form with $1/n$ and $1/j^2$ (and $1/n^3$) replaced by $1/\beta_n$ and $1/\beta_j^2$ (and $(\beta_n - \beta_{n-1})/\beta_n^3$), respectively, in Theorem 1b above, where $0 < \beta_n \uparrow \infty$.

The following theorem establishes validicity of the strong law of large numbers under some mixing conditions. Part a) comprehends Kolmogorov's classic strong law of large numbers for independent identically distributed integrable real random variables and, as this law, can be generalized to the case of random variables with values in a real separable Banach space.

We shall use the ρ-mixing and the α-mixing concept of dependence of random variables. Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of random variables on a probability space $(\Omega,\mathscr{A},P)$. $\mathscr{F}_m^n$ denotes the σ-algebra generated by $(Z_m,\ldots,Z_n)$ for $m \le n$. The ρ-mixing and α-mixing coefficients are defined by

$$\rho_n := \sup_{k\in\mathbb{N}}\sup\{|corr(U,V)|; U \in L_2(\mathscr{F}_1^k), V \in L_2(\mathscr{F}_{k+n}^\infty), U,V \text{ real-valued}\},$$

$$\alpha_n := \sup_{k\in\mathbb{N}}\sup\{|P(A\cap B) - P(A)P(B)|; A \in \mathscr{F}_1^k, B \in \mathscr{F}_{k+n}^\infty\},$$

respectively. The sequence (Z_n) is called ρ-mixing, if $\rho_n \to 0$, and α-mixing, if $\alpha_n \to 0$ $(n \to \infty)$. It holds $4\alpha_n \le \rho_n$ (see, e.g., Györfi et al. (1989), p. 9, and Doukhan (1994), p. 4). $\log_+$ below denotes the nonnegative part of log, i.e., $\max\{0,\log\}$.

Theorem 2. *Let the real random variables Y_n be identically distributed.*

a) If $\mathbf{E}|Y_1| < \infty$ and if (Y_n) is independent or, more generally, $\rho-$mixing with

$$\sum_{n=1}^{\infty}\frac{1}{n}\rho_n < \infty,$$

$$\textit{e.g., if } \rho_n = O\left(\frac{1}{(\log n)^{1+\delta}}\right) \textit{ for some } \delta > 0,$$

then

$$\frac{1}{n}\sum_{i=1}^{n}Y_i \to \mathbf{E}Y_1 \quad a.s.$$

a_1) If $\mathbf{E}\{|Y_1|\log_+|Y_1|\} < \infty$ and if (Y_n) is $\rho-$mixing, then

$$\frac{1}{n}\sum_{i=1}^{n}Y_i \to \mathbf{E}Y_1 \quad a.s.$$

b) If $\mathbf{E}\{|Y_1|\log_+|Y_1|\} < \infty$ and if (Y_n) is $\alpha-$mixing with $\alpha_n = O(n^{-\alpha})$ for some $\alpha > 0$, then

$$\frac{1}{n}\sum_{i=1}^{n}Y_i \to \mathbf{E}Y_1 \quad a.s.$$

Proof. Let $Y_n \ge 0$ without loss of generality. We set $Y_n^{[c]} := Y_n 1_{\{Y_n \le c\}}$, $c > 0$.

a) We use a well-known truncation argument. Because of $\mathbf{E}Y_1 < \infty$, we have a.s. $Y_n = Y_n^{[n]}$ for some random index on. Therefore and because of $\mathbf{E}Y_n^{[n]} \to \mathbf{E}Y$, it suffices to show

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i^{[i]} - \mathbf{E}Y_i^{[i]}\right) \to 0 \quad a.s.$$

Because of $Y_n^{[n]} \geq 0$, $\mathbf{E}Y_n^{[n]} \leq \mathbf{E}Y < \infty$, by Theorem 1a it is enough to show

$$\sum_{n=1}^{\infty}\frac{1}{n^3}\mathbf{Var}\left\{\sum_{i=1}^{n} Y_i^{[i]}\right\} < \infty.$$

Application of Lemma 3a below for real random variables yields

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Y_i^{[i]}\right\} \leq Cn\mathbf{E}\left\{\left(Y_n^{[n]}\right)^2\right\} = Cn\mathbf{E}\left\{\left(Y_1^{[n]}\right)^2\right\}.$$

with some constant $C \in \mathbb{R}_+$. In the special case of independence one immediately obtains the inequality with $C = 1$. ¿From this and the well-known relation

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{n^2}\mathbf{E}\left\{\left(Y_1^{[n]}\right)^2\right\} &= \sum_{n=1}^{\infty}\sum_{i=1}^{n}\frac{1}{n^2}\int_{(i-1,i]} t^2 P_{Y_1}(dt) \\ &= \sum_{i=1}^{\infty}\int_{(i-1,i]} t^2 P_{Y_1}(dt)\sum_{n=i}^{\infty}\frac{1}{n^2} \\ &\leq \sum_{i=1}^{\infty}\frac{2}{i}\int_{(i-1,i]} t^2 P_{Y_1}(dt) \\ &\leq 2\mathbf{E}Y_1 < \infty\end{aligned}$$

we obtain the assertion.

a$_1$) Let $\varepsilon = \frac{1}{4}$, $\kappa = \frac{1}{8}$. From the integrability assumption we obtain as in the first step of the proof of Theorem 3 below (specialization to $X_n = const$) that

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_i^{[i^\kappa]}\right) \to 0 \quad a.s.$$

As in part a) it is enough to show

$$\sum_{n=1}^{\infty}\frac{1}{n^3}\mathbf{Var}\left\{\sum_{i=1}^{n} Y_i^{[i^\kappa]}\right\} < \infty.$$

Application of Lemma 3a below for real random variables yields

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Y_i^{[i^\kappa]}\right\} \leq C(\varepsilon)n^{1+\varepsilon}\mathbf{E}\left\{\left(Y_n^{[n^\kappa]}\right)^2\right\} \leq C(\varepsilon)n^{1+\varepsilon+2\kappa}$$

for some $C(\varepsilon) < \infty$ and thus the assertion.

b) Let $\kappa = \frac{1}{4}\min\{1,\alpha\}$. As in a$_1$) it is enough to show

$$\sum_{n=1}^{\infty}\frac{1}{n^3}\mathbf{Var}\left\{\sum_{i=1}^{n} Y_i^{[i^\kappa]}\right\} < \infty.$$

Application of Lemma 3b below for real random variables yields

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Y_i^{[i^\kappa]}\right\} = O\left(n^{2\kappa+2-\min\{1,\alpha\}}\log(n+1)\right)$$

and thus the assertion.

□

Part a) of the following lemma is due to Peligrad (1992), Proposition 3.7 and Remark 3.8. Part b) is an immediate consequence of an inequality of Dehling and Philipp (1982), Lemma 2.2. Parts c) and d) are due to Liebscher (1996), Lemma 2.1, and Rio (1993), pp. 592, 593, respectively.

Lemma 3. *a) Let (Z_n) be a $\rho-$mixing sequence of square integrable variables with values in a real separable Hilbert space. Then for each $\varepsilon > 0$*

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Z_i\right\} \le C(\varepsilon)n^{1+\varepsilon} \max_{i=1,\dots,n} \mathbf{Var} Z_i$$

for some $C(\varepsilon) < \infty$. If additionally

$$\sum_{n=1}^{\infty} \rho_{2^n} < \infty \text{ or, equivalently, } \sum_{n=1}^{\infty} \frac{1}{n}\rho_n < \infty,$$

then

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Z_i\right\} \le Cn \max_{i=1,\dots,n} \mathbf{Var} Z_i$$

for some $C < \infty$.

b) Let (Z_n) be an α-mixing sequence of essentially bounded random variables with values in a real separable Hilbert space with $\alpha_n = O(n^{-\alpha})$ for some $\alpha > 0$, then

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Z_i\right\} \le Cn^{2-\min\{1,\alpha\}} \log(n+1) \max_{i=1,\dots,n} (\text{ess sup } \|Z_i\|)^2$$

for some $C < \infty$. In the case $\alpha \neq 1$ the assertion holds without the factor $\log(n+1)$.

c) Let (Z_n) be an $\alpha-$mixing sequence of real random variables with $\alpha_n = O(n^{-\alpha})$ for some $\alpha > 1$ and $\mathbf{E}\{|Z_n|^{2\alpha/(\alpha-1)}\} < \infty$, $n \in \mathbb{N}$. Then

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Z_i\right\} \le Cn\log(n+1) \max_{i=1,\dots,n} \left(\mathbf{E}\left\{|Z_i|^{2\alpha/(\alpha-1)}\right\}\right)^{\frac{\alpha-1}{\alpha}}$$

for some $C < \infty$.

d) Let (Z_n) be a weakly stationary α-mixing sequence of identically distributed real random variables with $\alpha_n = O(\delta^n)$ for some $\delta \in (0,1)$ and $\mathbf{E}\{Z_1^2 \log_+ |Z_1|\} < \infty$, then

$$\mathbf{Var}\left\{\sum_{i=1}^{n} Z_i\right\} \le Cn\mathbf{E}\left\{Z_1^2 \log_+ |Z_1|\right\}$$

for some $C < \infty$.

3 Pointwise consistent regression estimates

In regression analysis, on the basis of an observed d-dimensional random predictor vector X one wants to estimate the non-observed real random response variable Y by $f(X)$ with a suitable measurable function

$f:\mathbb{R}^d \to \mathbb{R}$. In case of a square integrable Y one is often interested to minimize the L_2 risk or mean squared error $\mathbf{E}\{|f(X)-Y|^2\}$. As is well known the optimal f is then given by the regression function m of Y on X defined by $m(x) := \mathbf{E}\{Y|X=x\}$. This follows from

$$\mathbf{E}\left\{|f(X)-Y|^2\right\} = \int_{\mathbb{R}^d} |f(x)-m(x)|^2 \mu(dx) + \mathbf{E}\left\{|m(X)-Y|^2\right\},$$

where μ denotes the distribution of X. Usually the distribution $P_{(X,Y)}$ of (X,Y), especially m, is unknown. If there is the possibility to observe a training sequence $(X_1,Y_1),(X_2,Y_2),\ldots$ of $(d+1)$-dimensional random vectors distributed like (X,Y) up to the index n, one now wants to estimate m by $m_n(x) := m_n(X_1,Y_1,\ldots,X_n,Y_n;x)$ in such a way that

$$\int |m_n(x)-m(x)|^2 \mu(dx) \to 0 \quad (n \to \infty)$$

almost surely (a.s.) or at least in probability. Inspired by $m(x) = \mathbf{E}(Y|X=x)$, $x \in \mathbb{R}^d$, one uses local averaging methods, where $m(x)$ is estimated by the average of those Y_i where X_i is close to x. Inspired by the above minimum property one also uses least squares methods, which minimize the empirical L_2 risk over a general set $\mathscr{F}_n$ of functions. The classic partitioning regression estimate (regressogram) is a local averaging method as well as a least squares method where $\mathscr{F}_n$ consists of the functions which are constant on each set belonging to a partition $\mathscr{P}_n$ of $\mathbb{R}^d$.
A frequently used local averaging estimate is the regression kernel estimate of Nadaraya and Watson. It uses a kernel function $K:\mathbb{R}^d \to \mathbb{R}_+$, usually with $0 < \int K(x)\lambda(dx) < \infty$ (λ denoting the Lebesgue-Borel measure on $\mathscr{B}_d$), e.g., $K = 1_{S_{0,1}}$ (naive kernel), $K(x) = (1-\|x\|^2)1_{S_{0,1}}(x)$ (Epanechnikov kernel), $K(x) = (1-\|x\|^2)^2 1_{S_{0,1}}(x)$ (quartic kernel) and $K(x) = e^{-\|x\|^2/2}$ (Gaussian kernel), with $x \in \mathbb{R}^d$, and bandwidth $h_n \in (0,\infty)$, usually satisfying $h_n \to 0$, $nh_n^d \to \infty$ $(n \to \infty)$, e.g., $h_n = cn^{-\gamma}$ $(c>0,\ 0<\gamma d<1)$. ($S_{x,h}$ for $x \in \mathbb{R}^d$, $h>0$ denotes the sphere in $\mathbb{R}^d$ with center x and radius h.) The estimator m_n is defined by

$$m_n(x) := \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}, \quad x \in \mathbb{R}^d \tag{10}$$

with $0/0 := 0$. The k_n-nearest neighbor ($k_n - NN$) regression estimate m_n of m is defined by

$$m_n(x) := \frac{1}{k_n}\sum_{i=1}^n Y_i 1_{\{X_i \text{ is among the } k_n \text{ NNs of } x \text{ in } (X_1,\ldots,X_n)\}} \tag{11}$$

with $k_n \in \{1,\ldots,n-1\}$, $n \geq 2$, usually satisfying $k_n/n \to 0$, $k_n \to \infty$ $(n \to \infty)$. Ambiguities in the definition of NNs (on the basis of the Euclidean distance in $\mathbb{R}^d$) can be solved by random tie-breaking. As to references see Györfi et a. (2002).

A regression estimation sequence is called strongly universally (L_2-)consistent (usually in the case that the sequence of identically distributed $(d+1)$-dimensional random vectors $(X_1,Y_1),(X_2,Y_2),\ldots$ is independent), if

$$\int |m_n(x)-m(x)|^2 \mu(dx) \to 0 \quad a.s. \tag{12}$$

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

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

for all distributions of (X,Y) with $\mathbf{E}|Y| < \infty$. (modμ means that the assertion holds for μ-almost all $x \in \mathbb{R}^d$.) Correspondingly one speaks of weak consistency if one has convergence in first mean (or in probability).

Results on strong universal pointwise or L_2-consistency will be stated which generalize Kolmogorov's strong law of large numbers. If the independence condition there is relaxed to a mixing condition, mostly the moment condition $\mathbf{E}|Y| < \infty$ or $\mathbf{E}\{|Y|^2\} < \infty$ for pointwise or L_2-consistency, respectively, has to be strengthened to $\mathbf{E}\{|Y|\log_+|Y|\} < \infty$ or higher moment conditions. We shall use ρ-mixing and α-mixing conditions. No continuity assumptions on the distribution of X will be made.
This section and the next section deal with strong pointwise consistency and with strong L_2-consistency, respectively.
In this section, more precisely, strong pointwise consistency of Nadaraya-Watson estimates (Theorem 3), strong universal pointwise consistency of semi-recursive Devroye-Wagner estimates (Theorem 4), both under mixing conditions, and strong universal pointwise consistency of k_n-nearest neighbor estimates under independence (Theorem 5) are stated.

Theorem 3. *Let $(X,Y),(X_1,Y_1),(X_2,Y_2),\ldots$ be identically distributed $(d+1)-$dimensional random vectors with $\mathbf{E}\left\{|Y|\log_+|Y|\right\} < \infty$. Let K be a measurable function on $\mathbb{R}^d$ satisfying $c_1H(\|x\|) \le K(x) \le c_2H(\|x\|)$, $x \in \mathbb{R}^d$, for some $0 < c_1 < c_2 < \infty$ and a nondecreasing function $H : \mathbb{R}_+ \to \mathbb{R}_+$ with $H(+0) > 0$ and $t^dH(t) \to 0$ $(t \to \infty)$, e.g., naive, Epanechnikov, quartic and Gaussian kernel. For $n \in \mathbb{N}$, with bandwidth $h_n > 0$, define m_n by (10).*

a) If the sequence $((X_n,Y_n))$ is $\rho-$mixing with $\rho_n = O(n^{-\rho})$ for some $\rho > 0$ and if h_n is chosen as $h_n = cn^{-\gamma}$ with $c > 0$, $0 < \gamma d < 2\rho/(1+2\rho)$, then

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

b) If the sequence $((X_n,Y_n))$ is $\alpha-$mixing with $\alpha_n = O(n^{-\alpha})$ for some $\alpha > 1$ and if h_n is chosen as $h_n = cn^{-\gamma}$ with $c > 0$, $0 < \gamma d < (2\alpha-2)/(2\alpha+3)$, then

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

Remark 4. Theorem 3 in both versions a) and b) comprehends the case of independent identically distributed random vectors with choice $h_n = cn^{-\gamma}$ satisfying $0 < \gamma d < 1$ treated in Kozek, Leslie and Schuster (1998), Theorem 2, with a somewhat more general choice of h_n, but with a somewhat stronger integrability condition such as $\mathbf{E}\left\{|Y|\log_+|Y|(\log_+\log_+|Y|)^{1+\delta}\right\} < \infty$ for some $\delta > 1$. In the proof of Theorem 3 exponential inequalities of Peligrad (1992) and Rhomari (2002) together with the above variance inequalities and a generalized Lebesgue density theorem due to Greblicki, Krzyżak and Pawlak (1984) together with a covering lemma for kernels are used. In the independence case the classic Bernstein exponential inequality, see Györfi et al. (2002), Lemma A.2, can be used.

Regarding Lemma 3a,c we can state the exponential inequalities of Peligrad (1992) and Rhomari (2002) for ρ-mixing and α-mixing sequences, respectively, of bounded real random variables in the following somewhat specialized form.

Lemma 4. *Let Z_n, $n \in \mathbb{N}$, be bounded real random variables and set*

$$L_n := \max_{i=1,\ldots,n} \text{ess sup } |Z_i|.$$

a) Let (Z_n) be ρ-mixing with $\rho_n = O(n^{-\rho})$ for some $\rho > 0$. Then there are constants $c_1, c_2 \in (0,\infty)$ such that for all $n \in \mathbb{N}$, $\varepsilon^* > 0$

$$P\left\{\left|\sum_{i=1}^{n}(Z_i - \mathbf{E}Z_i)\right| > \varepsilon^*\right\} \le c_1 \exp\left(-\frac{c_2\varepsilon^*}{n^{1/2}\max_{i=1,\ldots,n}(\mathbf{E}\{|Z_i|^2\})^{1/2} + L_n n^{1/(1+2\rho)}}\right).$$

b) Let (Z_n) be α-mixing with $\alpha_n = O(n^{-\alpha})$ for some $\alpha > 1$. Then there are constants $c_1, c_2 \in (0,\infty)$ such that for all $n \in \mathbb{N}$, $\varepsilon^* > 0$, $\beta \in (0,1)$

$$P\left\{\left|\sum_{i=1}^{n}(Z_i-\mathbf{E}Z_i)\right|>\varepsilon^*\right\}$$
$$\leq 4\exp\left(-\frac{c_1(\varepsilon^*)^2}{n\log(n+1)\max_{i=1,\dots,n}(\mathbf{E}\{|Z_i|^{2\alpha/(\alpha-1)}\})^{(\alpha-1)/\alpha}+\varepsilon^* L_n n^\beta}\right)$$
$$+c_2\max\left\{\left(\frac{L_n n}{\varepsilon^*}\right)^{\frac{1}{2}},1\right\}n^{1-\beta-\beta\alpha}.$$

The following generalized Lebesgue density theorem is due to Greblicki, Krzyżak and Pawlak (1984) (see also Györfi et al. (2002), Lemma 24.8).

Lemma 5. *Let K as in Theorem 3, $0<h_n\to 0$ $(n\to\infty)$, and let μ be a probability measure on $\mathscr{B}_d$. Then for all μ-integrable functions $f:\mathbb{R}^d\to\mathbb{R}$,*

$$\frac{\int K\left(\frac{x-t}{h_n}\right)f(t)\mu(dt)}{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}\to f(x)\quad \bmod \mu.$$

The next lemma is due to Devroye (1981) (see also Györfi et al. (2002), Lemma 24.6).

Lemma 6. *Let μ be a probability mesaure on $\mathscr{B}_d$ and $0<h_n\to 0$ $(n\to\infty)$. Then*

$$\liminf\frac{\mu(S_x,h_n)}{h_n^d}>0\quad \bmod \mu.$$

It follows a covering lemma. It can be proven as Lemma 23.6 in Györfi et al. (2002) where $K=\widetilde{K}$ is used.

Lemma 7. *Let $H,\widetilde{H}$ and $K,\widetilde{K}$ be functions as H and K, respectively, in Theorem 3. There exists $\tilde{c}\in(0,\infty)$ depending only on K and $\widetilde{K}$ such that for all $h>0$ and $u\in\mathbb{R}^d$*

$$\int\frac{\widetilde{K}\left(\frac{x-u}{h}\right)}{\int K\left(\frac{x-t}{h}\right)\mu(dt)}\mu(dx)\leq\tilde{c}.$$

Proof (of Theorem 3). It suffices to show

$$\overline{m}_n(x):=\frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_1}{h_n}\right)}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}\to m(x)\quad a.s.\quad \bmod \mu, \tag{13}$$

because this result together with its special case for $Y_i=\text{const}=1$ yields the assertion. Let $Y_i\geq 0$, $0\leq K\leq 1$, without loss of generality.

In the first step, for an arbitrary fixed $\kappa>0$ and $Y_i^*:=Y_i^{[i^\kappa]}:=Y_i 1_{[Y_i\leq i^\kappa]}$, we show

$$\frac{\sum_{i=1}^n (Y_i-Y_i^*)K\left(\frac{x-X_i}{h_n}\right)}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}\to 0\quad a.s.\quad \bmod \mu, \tag{14}$$

which together with (16) below yields the assertion. The notation $K_h(x)$ for $K(\frac{x}{h})$ $(x\in\mathbb{R}^d,\ h>0)$ will be used.

According to a monotonicity argument of Etemadi (1981), for (14) it suffices to show

$$V_n(x):=\frac{\sum_{i=1}^{2^{n+1}}(Y_i-Y_i^*)K_{h_{2^n}}(x-X_i)}{2^n\int K_{h_{2^{n+1}}}(x-t)\mu(dt)}\to 0\quad a.s.\quad \bmod \mu.$$

We notice

$$h_{2^n}/h_{2^{n+1}} = 2^{\gamma},$$

thus

$$K_{h_{2^n}} = K_{h_{2^{n+1}}}\left(\frac{\cdot}{2^{\gamma}}\right) =: \widetilde{K}_{h_{2^{n+1}}}$$

and, because of Lemma 7,

$$\int \frac{K_{h_{2^n}}(x-z)}{\int K_{h_{2^{n+1}}}(x-t)\mu(dt)}\mu(dx)$$
$$= \int \frac{\widetilde{K}_{h_{2^{n+1}}}(x-z)}{\int K_{h_{2^{n+1}}}(x-t)\mu(dt)}\mu(dx) \le \tilde{c} < \infty$$

for all $z \in \mathbb{R}^d$ and all n. Therefore, with suitable constants c_3, $c_4(\kappa)$,

$$\begin{aligned}
\mathbf{E}\sum_{n=1}^{\infty}\int V_n(x)\mu(dx) &\le \tilde{c}\sum_{n=1}^{\infty}2^{-n}\sum_{i=1}^{2^{n+1}}\mathbf{E}\{Y_i - Y_i^*\}\\
&\le \tilde{c}\sum_{i=1}^{\infty}\left(\sum_{n=\max\{1,\lfloor(\log i)/(\log 2)\rfloor-1\}}^{\infty}2^{-n}\right)\mathbf{E}Y_i 1_{[Y_i>i^{\kappa}]}\\
&\le c_3\sum_{i=1}^{\infty}\frac{1}{i}\sum_{j=\lfloor i^{\kappa}\rfloor}^{\infty}\int_{(j,j+1]}vP_Y(dv)\\
&\le c_3\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\lceil (j+1)^{\frac{1}{\kappa}}\rceil}\frac{1}{i}\right)\int_{(j,j+1]}vP_Y(dv)\\
&\le c_4(\kappa)\mathbf{E}\left\{Y\log_+ Y\right\} < \infty.
\end{aligned}$$

This yields (14). − In the second step we show

$$\frac{\sum_{i=1}^n \mathbf{E}\left\{Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right\}}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} \to m(x) \bmod \mu. \tag{15}$$

We have

$$\frac{\sum_{i=1}^n \mathbf{E}\left\{Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right\}}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} \le \frac{\mathbf{E}\left\{YK(\frac{x-X}{h_n})\right\}}{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}$$
$$= \frac{\int m(t)K\left(\frac{x-t}{h_n}\right)\mu(dt)}{\int K\left(\frac{x-t}{h_n}\right)\mu(dx)} \to m(x) \bmod \mu$$

by Lemma 5. Because of Lemma 6 we have

$$n\int K\left(\frac{x-t}{h_n}\right)\mu(dt) \ge d^*(x)n^{1-\gamma d} \to \infty \bmod \mu$$

(compare (18) below), thus

$$\liminf_{n\to\infty} \frac{\sum_{i=1}^n \mathbf{E}\left\{Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right\}}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} \geq \lim_{n\to\infty} \frac{n\mathbf{E}\left\{Y 1_{[Y\leq N]} K\left(\frac{x-X}{h_n}\right)\right\}}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}$$

$$= \lim_{n\to\infty} \frac{\int \mathbf{E}\left\{Y 1_{[Y\leq N]} \mid X=t\right\} K\left(\frac{x-t}{h_n}\right)\mu(dt)}{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}$$

$$= \mathbf{E}(Y 1_{[Y\leq N]} \mid X = x) \bmod \mu$$

(for each $N \in \mathbb{N}$, by Lemma 5).

$$\to \mathbf{E}(Y \mid X = x) = m(x) \quad (N\to\infty),$$

which leads to (15). Together with (17) below we shall have

$$\frac{\sum_{i=1}^n Y_i^* K\left(\frac{x-X_i}{h_n}\right)}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} \to m(x) \quad a.s. \quad \bmod \mu, \tag{16}$$

which together with (14) yields (13). − In the third step we show

$$\frac{\sum_{i=1}^n \left[Y_i^* K\left(\frac{x-X_i}{h_n}\right) - \mathbf{E}Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right]}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} \to 0 \quad a.s. \quad \bmod \mu \tag{17}$$

distinguishing the cases of a) ρ-mixing and b) α-mixing.

a) According to Lemma 6

$$\mu\left(\{x\in\mathbb{R}^d; \liminf \frac{\mu(S_{x,h_n})}{h_n^d} = 0\}\right) = 0.$$

Neglecting this set we have

$$\int K\left(\frac{x-t}{h_n}\right)\mu(dt) \geq c^* \int 1_{S_{0,r^*}}\left(\frac{x-t}{h_n}\right)\mu(dt)$$
$$\geq d(x)h_n^d = d^*(x) n^{-\gamma d} \tag{18}$$

for all n with suitable $c^* > 0$, $r^* > 0$, $d^*(x) > 0$. Choose an arbitrary $\varepsilon > 0$. Noticing

$$\mathbf{E}\left\{\left(Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right)^2\right\} \leq n^{2\kappa} \int K\left(\frac{x-t}{h_n}\right)\mu(dt) \quad (i=1,\ldots,n),$$

by Lemma 4a with $\varepsilon^* = \varepsilon n \int K(\frac{x-t}{h_n})\mu(dt)$ we obtain for suitable $c_1, c_2 \in (0,\infty)$

$$
\begin{aligned}
&\sum_{n=1}^{\infty} P\left\{\frac{1}{n\int K(\frac{x-t}{h_n})\mu(dt)}\left|\sum_{i=1}^{n}\left[Y_i^* K\left(\frac{x-X_i}{h_n}\right)-\mathbf{E}Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right]\right|>\varepsilon\right\}\\
&\le c_1\sum_{n=1}^{\infty}\exp\left(-\frac{c_2\varepsilon n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}{n^{1/2}n^{\kappa}(\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\Big)^{1/2}+n^{\kappa}n^{1/(1+2\rho)}}\right)\\
&\le c_1\sum_{n=1}^{\infty}\exp\Bigg(-\frac{1}{2}c_2\varepsilon\\
&\quad\cdot\min\left\{n^{\frac{1}{2}-\kappa}\left(\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\right)^{1/2},\,n^{1-\kappa-\frac{1}{1+2\rho}}\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\right\}\Bigg)\\
&\le c_1\sum_{n=1}^{\infty}\exp\left(-\frac{1}{2}c_2\varepsilon\min\left\{d^*(x)^{\frac{1}{2}}n^{\frac{1}{2}-\kappa-\frac{1}{2}\gamma d},d^*(x)n^{1-\kappa-\frac{1}{1+2\rho}-\gamma d}\right\}\right)\\
&\quad\text{(by 18)}\\
&=c_1\sum_{n=1}^{\infty}\exp\left(-\frac{1}{2}c_2\varepsilon\min\left\{d^*(x)^{\frac{1}{2}},d^*(x)\right\}n^{\min\left\{\frac{1}{2}-\kappa-\frac{1}{2}\gamma d,1-\kappa-\frac{1}{1+2\rho}-\gamma d\right\}}\right)\\
&<\infty \bmod \mu,
\end{aligned}
$$

if $1-\gamma d-2\kappa>0$ and $1-1/(1+2\rho)-\gamma d-\kappa>0$. Both conditions are fufilled under the assumptions on ρ and γ, if $\kappa>0$ is chosen sufficiently small. Thus (17) is obtained by the Borel-Cantelli lemma.

b) As in a) we have (18). Choose an arbitrary $\varepsilon>0$. For suitable constants c_1, c_2 by Lemma 4b with $\varepsilon^*=\varepsilon n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)$ and $\beta\in(0,1)$ we obtain for ε sufficiently small

$$
\sum_{n=1}^{\infty} P\left\{\frac{1}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}\left|\sum_{i=1}^{n}\left[Y_i^* K\left(\frac{x-X_i}{h_n}\right)-\mathbf{E}Y_i^* K\left(\frac{x-X_i}{h_n}\right)\right]\right|>\varepsilon\right\}
$$

$$
\begin{aligned}
&\le 4\sum_{n=1}^{\infty}\exp\left(\frac{-c_1\varepsilon^2 n^2\left(\int K(\frac{x-t}{h_n})\mu(dt)\right)^2}{n^{2\kappa+1}\log(n+1)(\int K(\frac{x-t}{h_n})\mu(dt))^{1-1/\alpha}+\varepsilon n^{1+\kappa+\beta}\int K(\frac{x-t}{h_n})\mu(dt)}\right)\\
&\quad+c_2\left(\frac{n^{\kappa}}{\varepsilon\int K(\frac{x-t}{h_n})\mu(dt)}\right)^{\frac{1}{2}}n^{1-\beta-\beta\alpha}\\
&\le 4\sum_{n=1}^{\infty}\exp\Bigg(-\frac{c_1\varepsilon}{2}\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\\
&\quad\cdot\min\left\{\varepsilon n^{1-2\kappa}(\log(n+1))^{-1}\left(\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\right)^{1/\alpha},\ n^{1-\kappa-\beta}\right\}\Bigg)\\
&\quad+c_2\varepsilon^{-\frac{1}{2}}d^*(x)^{-\frac{1}{2}}\sum_{n=1}^{\infty}n^{\frac{\kappa}{2}+\frac{\gamma d}{2}+1-(1+\alpha)\beta}\\
&\le 4\sum_{n=1}^{\infty}\exp\Bigg(-\frac{c_1\varepsilon}{2}d^*(x)n^{-\gamma d}\\
&\quad\cdot\min\left\{\varepsilon d^*(x)^{1/\alpha}n^{1-2\kappa-\gamma d/\alpha}(\log(n+1))^{-1},\ n^{1-\kappa-\beta}\right\}\Bigg)\\
&\quad+c_2\varepsilon^{-\frac{1}{2}}d^*(x)^{-\frac{1}{2}}\sum_{n=1}^{\infty}n^{\frac{\kappa}{2}+\frac{\gamma d}{2}+1-(1+\alpha)\beta}<\infty \bmod \mu,
\end{aligned}
$$

if $1-\gamma d(\alpha+1)/\alpha-2\kappa>0$, $1-\beta-\gamma d-\kappa>0$ and $-4+2(1+\alpha)\beta-\kappa-\gamma d$ >0. These conditions are fulfilled under the assumptions on α and γ, if one chooses $\beta=5/(3+2\alpha)$ and $\kappa>0$ sufficiently small. Thus (17) is obtained. □

If the above Nadaraya-Watson kernel regression estimate is replaced by the semi-recursive Devroye-Wagner (1980b) kernel regression estimate, then strong universal pointwise consistency in the case of independent identically distributed random vectors (X_n,Y_n) can be stated, i.e., under the only condition $\mathbf{E}|Y_1|<\infty$ strong consistency P_{X_1}-almost everywhere (see Walk (2001)). This result has been extended to the ρ-mixing case under the condition $\sum\rho_n<\infty$ by Frey (2007). The case of bounded Y was treated by Ferrario (2004) under more general α-mixing and ρ-mixing conditons on the basis of the generalized Theorem 1b mentioned in context of Lemma 2.
In the following the result of Frey (2007) and its proof will be given.

Theorem 4. *Let (X,Y), (X_1,Y_1), $(X_2,Y_2),\ldots$ be identically distributed $(d+1)$-dimensional random vectors with $\mathbf{E}|Y|<\infty$. Let K be a symmetric measurable function on $\mathbb{R}^d$ satisfying $c_1 1_{S_{0,R}}\le K\le c_2 1_{S_{0,R}}$ for some $0<R<\infty$, $0<c_1<c_2<\infty$ (so-called boxed kernel with naive kernel $K=1_{S_{0,1}}$ as a special case). With $n\in\mathbb{N}$ and $h_n>0$ set*

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

where $\frac{0}{0}:=0$. If the sequence $((X_n,Y_n))$ is ρ-mixing with $\sum\rho_n<\infty$ (e.g., $\rho_n=O(n^{-\rho})$ for some $\rho>1$) and if h_n is choosen as $h_n=cn^{-\gamma}$ with $c>0$, $0<\gamma d<\frac{1}{2}$, then

$$m_n(x)\to m(x)\quad a.s.\quad \mathrm{mod}\ \mu.$$

Theorem 4 comprehends Kolmogorov's strong law of large numbers (special case that μ is a Dirac measure). The semirecursive kernel estimate has the numerical advantage that a new observation leads only to an addition of a new summand in the numerator and in the denominator, but the observations obtain different weights. In the proof we give in detail only the part which differs from the proof in Walk (2001).

Proof (of Theorem 4). Without loss of generality assume $Y_i\ge 0$.The case of bounded Y, also with denominator replaced by its expectation, is comprehended by Ferrario (2004). Therefore in the case $\mathbf{E}|Y|<\infty$ it is enough to show existence of a $c\in(0,\infty)$ independent of the distribution of (X,Y) with

$$\limsup_{n\to\infty}\frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_i}\right)}{1+\sum_{i=1}^n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}\le cm(x)\quad a.s.\quad \mathrm{mod}\ \mu \tag{19}$$

(compare Lemma 8 below). Let the compact support of K by covered by finitely many closed spheres $S_1,\ldots,S_N$ with radius $R/2$. Fix $k\in\{1,\ldots,N\}$. For all $t\in\mathbb{R}^d$ and all $n\in\mathbb{N}$, from $x\in t+h_nS_k$ it can be concluded

$$K\left(\frac{\cdot-x}{h_i}\right)\ge\frac{c_1}{c_2}K\left(\frac{\cdot-t}{h_i}\right)1_{S_k}\left(g\left(\frac{\cdot-t}{h_i}\right)\right) \tag{20}$$

for all $i=\{1,\ldots,n\}$. It suffices to show

$$\limsup_{n\to\infty}\frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_i}\right)1_{S_k}\left(\frac{x-X_i}{h_i}\right)}{1+\sum_{i=1}^n\int K\left(\frac{x-t}{h_i}\right)\mu(dt)}\le c'm(x)\quad a.s.\quad \mathrm{mod}\ \mu \tag{21}$$

for some $c'<\infty$. With

$$r_n:=r_n(t):=\frac{1}{c_2}\int K\left(\frac{x-t}{h_n}\right)1_{t+h_nS_k}(x)\mu(dx),\ R_n:=r_1+\cdots+r_n,\ n\in\mathbb{N},$$

for $t \in \mathbb{R}^d$ we can choose indices $p_i = p(t,k,i) \uparrow \infty$ $(i \to \infty)$ such that

$$R_{p_i} \le i+1, \tag{22}$$

$$\sum_{j=p_i}^{\infty} \frac{r_j}{(1+R_j)^2} < \frac{1}{i} \tag{23}$$

holds. For $p(t,k,\cdot)$ we define the inverse function $q(t,k,\cdot)$ on $\mathbb{N}$ by

$$q(t,k,n) := \max\{i \in \mathbb{N}; p(t,k,i) \le n\}.$$

Set

$$Z_i := Y_i 1_{[Y_i \le q(X_i,k,i)]}, \quad i \in \mathbb{N}.$$

Now it will be shown

$$\frac{\sum_{i=1}^n \left[Z_i K\left(\frac{x-X_i}{h_i}\right) 1_{S_k}\left(\frac{x-X_i}{h_i}\right) - \mathbf{E}\left\{ Z_i K\left(\frac{x-X_i}{h_i}\right) 1_{S_k}\left(\frac{x-X_i}{h_i}\right)\right\}\right]}{c_1 + \sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt)} \to 0 \quad a.s. \quad \mod \mu \tag{24}$$

by an application of Lemma 2. We notice $\int K(\frac{x-t}{h_n})\mu(dt) \le c_2$,

$$\sum_{i=1}^n \int K\left(\frac{x-t}{h_i}\right)\mu(dt) \uparrow \infty \quad \mod \mu \tag{25}$$

because of $\int K\left(\frac{x-t}{h_i}\right)\mu(dt) \ge c_1 \mu(x + h_i S_{0,1}) \ge c_1 c(x) h_i^d$ by Lemma 6 with $c(x) > 0 \mod \mu$ and $\sum h_n^d = \infty$ by $0 < \gamma d < 1$. Further $Z_n \ge 0$ and

$$\limsup \frac{\mathbf{E} Z_n K\left(\frac{x-X_n}{h_n}\right) 1_{S_k}\left(\frac{x-X_n}{h_n}\right)}{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} \le \lim \frac{\int m(t) K\left(\frac{x-t}{h_n}\right)\mu(dt)}{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} = m(x) \quad \mod \mu$$

by Lemma 5. With $W_j(x) := Z_j K\left(\frac{x-X_j}{h_j}\right) 1_{S_k}\left(\frac{x-X_j}{h_j}\right)$ we obtain

$$\begin{aligned}
&\mathbf{Var}\left\{\sum_{j=1}^n W_j(x)\right\} \\
&\le \sum_{j=1}^n \mathbf{Var}\{W_j(x)\} + \sum_{j=1}^n \sum_{l=1, l\ne j}^n \rho_{|j-l|} \left(\mathbf{Var}\{W_j(x)\}\right)^{\frac{1}{2}} \left(\mathbf{Var}\{W_l(x)\}\right)^{\frac{1}{2}} \\
&\le \sum_{j=1}^n \mathbf{Var}\{W_j(x)\} + \frac{1}{2}\sum_{j=1}^n \sum_{l=1, l\ne j}^n \rho_{|j-l|} \left[\mathbf{Var}\{W_j(x)\} + \mathbf{Var}\{W_l(x)\}\right] \\
&\le \left(1 + 2\sum_{j=1}^\infty \rho_j\right) \sum_{j=1}^n \mathbf{Var}\{W_j(x)\} \\
&= c^* \sum_{j=1}^n \mathbf{Var}\{W_j(x)\}
\end{aligned}$$

with $c^* < \infty$ by the assumption on (ρ_n), thus

$$\begin{aligned}
&\sum_{n=1}^{\infty} \frac{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\mathbf{Var}\left\{\sum_{j=1}^{n} W_j(x)\right\}}{\left(c_1+\sum_{i=1}^{n}\int K\left(\frac{x-t}{h_i}\right)\mu(dt)\}\right)^3} \\
&\le c^* \sum_{n=1}^{\infty} \frac{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)\sum_{j=1}^{n}\mathbf{Var}\left\{W_j(x)\right\}}{\left(c_1+\sum_{i=1}^{n}\int K\left(\frac{x-t}{h_i}\right)\mu(dt)\right)^3} \\
&= c^* \sum_{j=1}^{\infty}\mathbf{Var}\left\{W_j(x)\right\}\sum_{n=j}^{\infty} \frac{\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}{\left(c_1+\sum_{i=1}^{n}\int K\left(\frac{x-t}{h_i}\right)\mu(dt)\right)^3} \\
&\le c^{**} \sum_{j=1}^{\infty} \frac{\mathbf{Var}\left\{W_j(x)\right\}}{\left(c_1+\sum_{i=1}^{j}\int K\left(\frac{x-t}{h_i}\right)\mu(dt)\right)^2} \\
&\le c^{**}c_2 \sum_{n=1}^{\infty} \frac{\mathbf{E}Z_n^2 K\left(\frac{x-X_n}{h_n}\right)1_{S_k}\left(\frac{x-X_n}{h_n}\right)}{\left(c_1+\sum_{i=1}^{n}\int K\left(\frac{x-s}{h_i}\right)\mu(ds)\right)^2}
\end{aligned}$$

with suitable $c^{**} < \infty$. Now, by (20),

$$\begin{aligned}
&\int \sum_{n=1}^{\infty} \frac{\mathbf{E}Z_n^2 K\left(\frac{x-X_n}{h_n}\right)1_{S_k}\left(\frac{x-X_n}{h_n}\right)}{\left(c_1+\sum_{i=1}^{n}\int K\left(\frac{x-s}{h_i}\right)\mu(ds)\right)^2}\mu(dx) \\
&\le \sum_{n=1}^{\infty}\int\left[\int \frac{\mathbf{E}\left\{Z_n^2|X_n=t\right\}K\left(\frac{x-t}{h_n}\right)1_{S_k}\left(\frac{x-t}{h_n}\right)}{\left(c_1+\sum_{i=1}^{n}\int \frac{c_1}{c_2}K\left(\frac{s-t}{h_i}\right)1_{S_k}\left(\frac{s-t}{h_i}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt) \\
&= \frac{1}{c_1^2}\sum_{n=1}^{\infty}\int\left[\int \frac{\int v^2 P_{Z_n|X_n=t}(dv)K\left(\frac{x-t}{h_n}\right)1_{S_k}\left(\frac{x-t}{h_n}\right)}{\left(1+\sum_{i=1}^{n}\frac{1}{c_2}\int K\left(\frac{s-t}{h_i}\right)1_{S_k}\left(\frac{s-t}{h_i}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt) \\
&= \frac{c_2}{c_1^2}\sum_{n=1}^{\infty}\int\left[\int \frac{\sum_{i=1}^{q(t,k,n)}\left(\int_{(i-1,i]}v^2 P_{Z_n|X_n=t}(dv)\right)\frac{1}{c_2}K\left(\frac{x-t}{h_n}\right)1_{S_k}\left(\frac{x-t}{h_n}\right)}{\left(1+\sum_{i=1}^{n}\frac{1}{c_2}\int K\left(\frac{s-t}{h_i}\right)1_{S_k}\left(\frac{s-t}{h_i}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt) \\
&= \frac{c_2}{c_1^2}\sum_{n=1}^{\infty}\int\left[\int \frac{\sum_{i=1}^{q(t,k,n)}\left(\int_{(i-1,i]}v^2 P_{Y|X=t}(dv)\right)\frac{1}{c_2}K\left(\frac{x-t}{h_n}\right)1_{S_k}\left(\frac{x-t}{h_n}\right)}{\left(1+\sum_{i=1}^{n}\frac{1}{c_2}\int K\left(\frac{s-t}{h_i}\right)1_{S_k}\left(\frac{s-t}{h_i}\right)\mu(ds)\right)^2}\mu(dx)\right]\mu(dt) \\
&= \frac{c_2}{c_1^2}\int\left[\sum_{i=1}^{\infty}\int_{(i-1,i]}v^2 P_{Y|X=t}(dv)\right. \\
&\qquad \left.\sum_{n=p(t,k,i)}^{\infty} \frac{\frac{1}{c_2}\int K\left(\frac{x-t}{h_n}\right)1_{S_k}\left(\frac{x-t}{h_n}\right)\mu(dx)}{\left(1+\sum_{i=1}^{n}\frac{1}{c_2}\int K\left(\frac{s-t}{h_i}\right)1_{S_k}\left(\frac{s-t}{h_i}\right)\mu(ds)\right)^2}\right]\mu(dt)
\end{aligned}$$

$$\leq \frac{c_2}{c_1^2} \int \left[\sum_{i=1}^{\infty} \int_{(i-1,i]} v^2 P_{Y|X=t}(dv) \frac{1}{i} \right] \mu(dt)$$

(by 23)

$$\leq \frac{c_2}{c_1^2} \int \left[\sum_{i=1}^{\infty} \int_{(i-1,i]} v P_{Y|X=t}(dv) \right] \mu(dt)$$

$$\leq \frac{c_2}{c_1^2} \mathbf{E} Y < \infty.$$

Therefore

$$\sum_{n=1}^{\infty} \frac{\int K\left(\frac{x-t}{h_n}\right) \mu(dt) \mathbf{Var}\left\{\sum_{j=1}^{n} Z_j K\left(\frac{x-X_j}{h_j}\right) 1_{S_k}\left(\frac{x-X_j}{h_j}\right)\right\}}{\left(c_1 + \sum_{i=1}^{n} \int K\left(\frac{x-t}{h_i}\right) \mu(dt)\right)^3} < \infty \quad \mod \mu,$$

and (24) follows by Lemma 2.
In the next step we notice

$$\limsup \frac{\sum_{i=1}^{n} \mathbf{E} Z_i K\left(\frac{x-X_i}{h_i}\right) 1_{S_k}\left(\frac{x-X_i}{h_i}\right)}{c_1 + \sum_{i=1}^{n} \int K\left(\frac{x-t}{h_i}\right) \mu(dt)} \leq \lim \frac{\sum_{i=1}^{n} \int m(t) K\left(\frac{x-t}{h_i}\right) \mu(dt)}{c_1 + \sum_{i=1}^{n} \int K\left(\frac{x-t}{h_i}\right) \mu(dt)}$$
$$= m(x) \quad \mod \mu$$

because of (25) and Lemma 5. This together with (24) yields

$$\limsup \frac{\sum_{i=1}^{n} Z_i K\left(\frac{x-X_i}{h_i}\right) 1_{S_k}\left(\frac{x-X_i}{h_i}\right)}{c_1 + \sum_{i=1}^{n} \int K\left(\frac{x-t}{h_i}\right) \mu(dt)} \leq m(x) \quad a.s. \quad \mod \mu. \tag{26}$$

In the last step one obtains (21) from (26) and (25) by noticing

$$\sum_{n=1}^{\infty} P\left[Z_n 1_{S \cap S_k}\left(\frac{x-X_n}{h_n}\right) \neq Y_n 1_{S \cap S_k}\left(\frac{x-X_n}{h_n}\right) \right] < \infty \quad \mod \mu, \tag{27}$$

where $S := S_{0,R}$, together with the Borel-Cantelli lemma, and (27) follows from

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

$$\leq \frac{c_2}{c_1} \int \sum_{i=1}^{\infty} P\left[Y \in (i,i+1] | X = t\right] (i+2) \mu(dt)$$

(by 22)

$$\leq 3 \frac{c_2}{c_1} \mathbf{E} Y < \infty. \qquad \square$$

For k_n-nearest neighbor regression estimation with integrable response random variable Y and d-dimensional predictor random vector X on the basis of independent data, the following theorem states strong universal pointwise consistency, i.e., strong consistency P_X-almost everywhere for general distribution of (X,Y) with $\mathbf{E}|Y| < \infty$. The estimation is symmetric in the data, does not use truncated observations and contains Kolmogorov's strong law of large numbers as the special case that P_X is a Dirac measure. It can be considered as a universal strong law of large numbers for conditional expectations. Let for the observed copies of (X,Y) the k_n-nearest neighbor (k_n-NN) regression estimate $m_n(x)$ of $m(x) := \mathbf{E}(Y|X=x)$ be defined by (11).

Theorem 5. *Let (X,Y), (X_1,Y_1), $(X_2,Y_2),\dots$ be independent identically distributed $(d+1)$-dimensional random vectors with $\mathbf{E}|Y| < \infty$. Choose $k_n \in \min\{\lceil cn^\beta \rceil, n-1\}$ with $c > 0$, $\beta \in (0,1)$ for $n \in \{2,3,\dots\}$. Then*

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

As to the proof (with somewhat more general k_n) and related results we refer to Walk (2008a). The proof uses Etemadi's (1981) monotonicity argument, a generalized Lebesgue density theorem concerning $\mathbf{E}m_n(x) \to m(x) \bmod \mu$, a covering lemma for nearest neighbors and Steele's (1986) version of the Efron-Stein inequality for the variance of a function of independent identically distributed random variables.

Whether at least in the case of independence strong universal pointwise consistency for Nadaraya-Watson kernel regression estimates or for classic partitioning regression estimates holds, is an open problem.

Remark 5. Let the situation in Theorem 5 be modified by assuming that the sequence (X_1,Y_1), $(X_2,Y_2),\dots$ of identically distributed $(d+1)$-dimensional random vectors is α-mixing with $\alpha_n = O(n^{-\alpha})$ such that $0 < 1-\beta < \min\{\alpha/2, \alpha/(\alpha+1)\}$ and that Y_i is bounded. Let tie-breaking be done by enlarging the dimension d of the predictor vectors to $d+1$ by use of new (independent!) random variables equidistributed on $[0,1]$ as additional components (see Györfi et al. (2002), pp. 86,87). Then Theorem 2 in Irle (1997) states

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

Analogously, by use of Lemma 3a, one obtains the same convergence assertion under ρ-mixing where $0 < 1-\beta < 1$.

4 L_2-consistent regression estimates

The pioneering paper on universal consistency of nonparametric regression estimates is Stone (1977). It contains a criterion of weak universal L_2-consistency of local averaging estimates under independence. The conditions for k_n-nearest neighbor estimates and for Nadaraya-Watson kernel estimates were verified by Stone (1977) and by Devroye and Wagner (1980a) and Spiegelman and Sacks (1980), respectively. The following theorem concerns strong universal L_2-consistency of k_n-nearest neighbor estimates (Devroye et al. (1994)), of semirecursive Devroye-Wagner kernel estimates (Györfi, Kohler and Walk (1998)) and modified Nadaraya-Watson kernel estimates (Walk 2005a).

Theorem 6. *Let (X,Y), (X_1,Y_1), (X_2,Y_2), $\dots$ be independent identically distributed $(d+1)$-dimensional random vectors with $\mathbf{E}\{Y^2\} < \infty$.*

a) Let the k_n-NN regression estimates m_n of m be defined by (11) with $k_n \in \{1,\dots,n-1\}$, $n \geq 2$, satisfiying $k_n/n \to 0$, $k_n/\log n \to \infty$ $(n \to \infty)$ and random tie-breaking. Then (12) holds.
b) Let the semirecursive Devroye-Wagner kernel regression estimates m_n, $n \geq 2$, be defined by

$$m_n(x) := \frac{Y_1 K(0) + \sum_{i=2}^n Y_i K\left(\frac{x-X_i}{h_i}\right)}{K(0) + \sum_{i=2}^n K\left(\frac{x-X_i}{h_i}\right)}, \quad x \in \mathbb{R}^d,$$

with symmetric $\lambda-$integrable kernel $K:\mathbb{R}^d \to \mathbb{R}_+$ satisfying

$$\alpha H(||x||) \le K(x) \le \beta H(||x||), \quad x \in \mathbb{R}^d,$$

for some $0<\alpha<\beta<\infty$ and nonincreasing $H:\mathbb{R}_+ \to \mathbb{R}_+$ with $H(+0)>0$ and with bandwidths $h_n>0$ satisfying

$$h_n \downarrow 0 \quad (n\to\infty), \qquad \sum_{n=2}^{\infty} h_n^d = \infty,$$

e.g., $h_n = cn^{-\gamma}$ with $c>0$, $0<\gamma d<1$. Then (12) holds.

c) Let the Nadaraya-Watson type kernel regression estimates m_n, $n\in\mathbb{N}$, be defined by

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

with an arbitrary fixed $\delta>0$, a smooth kernel $K:\mathbb{R}^d\to\mathbb{R}_+$ (see below) and bandwidths $h_n>0$ satisfying

$$h_n \downarrow 0, \quad nh_n^d \to \infty \quad (n\to\infty), \quad h_n - h_{n+1} = O(h_n/n),$$

e.g., $h_n = cn^{-\gamma}$ with $c>0$, $0<\gamma d<1$. Then (12) holds.

In Theorem 6c the modification of Nadaraya-Watson estimates consists of a truncation of the denominator from below by an arbitrary positive constant, see Spiegelman and Sacks (1980)). Smooth kernel means a kernel K of the form $K(x)=H(||x||)$, where H is a continously differentiable nonincreasing function on $\mathbb{R}_+$ with $0<H(0)\le 1$, $\int H(s)s^{d-1}ds<\infty$ such that R with $R(s):=s^2H'(s)^2/H(s)$, $s\ge 0$ $(0/0:=0)$, is bounded, piecewise continuous and, for s sufficiently large, nonincreasing with $\int R(s)s^{d-1}ds<\infty$. Examples are the quartic and the Gaussian kernel. In the proof of Theorem 6a one shows

$$\limsup_{n\to\infty} \frac{n}{k_n} \max_{i=1,\dots,n} \int 1_{\{X_i \text{ is among the } k_n \text{ nearest neighbors of } x \text{ in } (X_1,\dots,X_n)\}}\mu(dx)$$
$$\le \text{const} < \infty \quad a.s.$$

and uses Kolmogorov's strong law of large numbers for $Y_1^2, Y_2^2, \dots$ and Lemma 8 (with $p=2,\ \delta=0$) below. Theorem 6b is proven by martingale theory, a covering argument and Lemmas 5, 6 and 8 (with $p=2,\ \delta=0$). In both cases, for details and further literature we refer to Györfi et al. (2002). In the proof of Theorem 6c strong consistency for bounded Y (due to Devroye und Krzyżak (1989)), Lemma 8 (with $p=2,\ \delta=0$) and summability theory (Lemma 1b), together with Lemmas 5, 6, 7 and Steele's (1986) version of the Efron-Stein inequality for variances are used.

The following lemma (see Györfi (1991), Theorem 2 with proof, and Györfi et al. (2002), Lemma 23.3; compare also the begin of the proof of Theorem 4) allows to reduce problems of strong consistency of kernel or nearest neighbor regression estimates to two simpler problems. It holds more generally for local averaging estimation methods.

Lemma 8. *Let $p\ge 1$ and $\delta\ge 0$ be fixed. Denote the Nadaraya-Watson or semirecursive Devroye-Wagner or k_n-NN regression estimates in the context of $(d+1)$-dimensional identically distributed random vectors $(X,Y),\ (X_1,Y_1),\ (X_2,Y_2),\dots$*
by m_n. The following statement a) is implied by statement b):

a) for all Y with $\mathbf{E}\left\{|Y|^{(1+\delta)p}\right\}<\infty$

$$\int |m_n(x)-m(x)|^p \mu(dx) \to 0 \quad a.s.;$$

b) for all bounded Y

$$\int |m_n(x)-m(x)|\mu(dx) \to 0 \quad a.s.,$$

and there exists a constant $c < \infty$ such that for all $Y \geq 0$ with $\mathbf{E}\left\{Y^{1+\delta}\right\} < \infty$

$$\limsup \int m_n(x)\mu(dx) \leq c\mathbf{E}Y \quad a.s. \tag{28}$$

For fixed $\delta \geq 0$ statement a) for $p > 1$ follows from a) for $p = 1$.

If we allow stronger moment conditions on Y (and X) we can relax the independence assumption for kernel estimates. Here Nadaraya-Watson kernel estimates m_n are considered with kernels $K : \mathbb{R}^d \to \mathbb{R}_+$ of the form $K(x) = H(\||x||)$, $x \in \mathbb{R}^d$, where $H : \mathbb{R}_+ \to \mathbb{R}_+$ is a Lipschitz continuous nonincreasing function with $0 < H(0) \leq 1$ and $\int H(s)s^{d-1}ds < \infty$ such that the function $s \to s|H'(s)|$ (defined λ-almost everywhere on $\mathbb{R}_+$) is nonincreasing for s sufficiently large (e.g., Epanechnikov, quartic and Gaussian kernel). The following result (Walk (2008b)) concerns L_2-consistency.

Theorem 7. *Let (X,Y), (X_1,Y_1), $(X_2,Y_2),\ldots$ be identically distributed $(d+1)$-dimensional random vectors with $\mathbf{E}\{|Y|^p\} < \infty$ for some $p > 4$, $\mathbf{E}\{||X||^q\} < \infty$ for some $q > 0$. Choose bandwidths $h_n = cn^{-\gamma}$ $(c > 0,\ 0 < \gamma d < 1)$. If the sequence $((X_n,Y_n))$ is ρ-mixing and $0 < \gamma d < 1 - \frac{4}{p} - \frac{2d}{pq}$ or if it is α-mixing with $\alpha_n = O(n^{-\alpha})$, $\alpha > 0$, with $0 < \gamma d < \min\{1,\alpha\} - \frac{4}{p} - \frac{2d}{pq}$, then*

$$\int |m_n(x) - m(x)|^2\mu(dx) \to 0 \quad a.s.$$

If Y is essentially bounded, then no moment condition on X is needed and the conditions on γ are $0 < \gamma d < 1$ in the ρ-mixing case and $0 < \gamma d < \min\{1,\alpha\}$ in the α-mixing case. If X is bounded, then the conditions on γ are $0 < \gamma d < 1 - \frac{4}{p}$ in the ρ-mixing case and $0 < \gamma d < \min\{1,\alpha\} - \frac{4}{p}$ in the α-mixing case. In this context we mention that a measurable transformation of X to bounded X does not change the L_2 risk $\mathbf{E}\{|Y - m(X)|^2\}$.
In the proof of Theorem 7, by Lemma 8 we treat the corresponding L_1-consistency problem with $p > 2$. The integrability assumption on Y allows to truncate Y_i (≥ 0) at $i^{1/p}$. Because of

$$\left|\frac{\sum_{i=1}^n Y_i 1_{\left[Y_i \leq i^{1/p}\right]} K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} - \frac{\sum_{i=1}^n Y_i 1_{\left[Y_i \leq i^{1/p}\right]} K\left(\frac{x-X_i}{h_n}\right)}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)}\right|$$
$$\leq n^{\frac{1}{p}} \left|\frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}{n\int K\left(\frac{x-t}{h_n}\right)\mu(dt)} - 1\right|,$$

it suffices to investigate the convergence behavior of the latter term and of the simplified estimator

$$\frac{\sum_{i=1}^n Y_i 1_{\left[Y_i \leq i^{1/p}\right]} K\left(\frac{\cdot-X_i}{h_n}\right)}{n\int K\left(\frac{\cdot-t}{h_n}\right)\mu(dt)}.$$

considered as a random variable with values in the real separable Hilbert space $L_2(\mu)$. This can be done by use of the Tauberian Lemma 1b, the covering Lemma 7 and Lemma 3a,b on the variance of sums of Hilbert space valued random variables under mixing together with a result of Serfling (1970a), Corollary A 3.1, on maximum cumulative sums.

5 Rosenblatt-Parzen density estimates under mixing conditions

In this section we investigate the Rosenblatt-Parzen kernel density estimates in view of strong L_1- consistency under mixing conditions, namely ρ- and α-mixing. In the latter case the α-mixing condition in Theorem 4.2.1 (iii) in Györfi et al. (1989) (see also Györfi and Masry (1990)) is weakened, essentially to that in Theorem 4.3.1 (iii) there on the Wolverton-Wagner and Yamato recursive density estimates. In the proof we use Etemadi's concept and not Tauberian theory, because in the latter case a Lipschitz condition on the kernel should be imposed.

Theorem 8. *Let the d-dimensional random vectors X_n, $n \in \mathbb{N}$, be identically distibuted with density f, and assume that (X_n) is ρ-mixing or α-mixing with $\alpha_n = O(n^{-\alpha})$ for some $\alpha > 0$. If for the Rosenblatt-Parzen density estimates*

$$f_n(x) := f_n(X_1,\ldots,X_n;x) := \frac{1}{nh_n^d}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{29}$$

the kernel K is chosen as a square λ-integrable density on $\mathbb{R}^d$ with $K(rx) \geq K(x)$ for $0 \leq r \leq 1$ and the bandwidths are of the form $h_n = cn^{-\gamma}$, $0 < c < \infty$, with $0 < \gamma d < 1$ in the ρ-mixing case and $0 < \gamma d <$ $\min\{1,\alpha\}$ in the α-mixing case, then

$$\int |f_n(x) - f(x)|\lambda(dx) \to 0 \quad a.s. \tag{30}$$

Proof. Because of the simple structure of the denominator in (29) we can use Etemadi's monotonicity argument. Let $(\Omega,\mathscr{A},P)$ be the underlying probability space. For rational $a > 1$ and $n \in \mathbb{N}$ set

$$q(a,n) := \min\{a^N; a^N > n, N \in \mathbb{N}\}, \quad p(a,n) := q(a,n)/a.$$

Then for f_n one has the majorant

$$g(a,n,\cdot) := \frac{1}{p(a,n)h_{q(a,n)}^d}\sum_{i=1}^{\lceil q(a,n)\rceil} K\left(\frac{\cdot - X_i}{h_{p(a,n)}}\right)$$

and a corresponding minorant $b(a,n,\cdot)$. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the norms in $L_1(\lambda)$ and $L_2(\lambda)$, respectively. In order to show

$$\left\| g(a,n,\cdot) - a^{1+\gamma d} f\right\|_1 \to 0 \quad (n \to \infty) \quad a.s.,$$

i.e.,

$$\left\| \sum_{i=1}^{\lceil a^{N+1}\rceil} V_{N,i} - a^{1+\gamma d} f\right\|_1 \to 0 \quad (N \to \infty) \quad a.s.,$$

with

$$V_{N,i}(x) := \frac{1}{a^N h_{a^{N+1}}^d} K\left(\frac{x-X_i}{h_{a^N}}\right),$$

it suffices, according to Györfi et al (1989), pp. 76, 77, to show

$$\left\| \sum_{i=1}^{\lceil a^{N+1}\rceil} \mathbf{E}V_{N,i} - a^{1+\gamma d} f\right\|_1 \to 0 \tag{31}$$

and

$$\left\| \sum_{i=1}^{\lceil a^{N+1}\rceil} (V_{N,i} - \mathbf{E}V_{N,i})\right\|_2 \to 0 \quad a.s. \tag{32}$$

(31) follows from Theorem 1 in Chapter 2 of Devroye and Györfi (1985). Noticing

$$\|V_{N,i}\|_2 = \frac{1}{a^N h_{a^{N+1}}^d} h_{a^N}^{\frac{d}{2}} \left(\int K(s)^2 \lambda(ds)\right)^{\frac{1}{2}},$$

one obtains

$$\sum_{N=1}^{\infty} \mathbf{E} \left\| \sum_{i=1}^{\lceil a^{N+1} \rceil} (V_{N,i} - \mathbf{E}V_{N,i}) \right\|_2^2 < \infty$$

by Lemma 3a,b and thus (32). Analogously one has

$$\|b(a,n,\cdot) - a^{-1-\gamma d} f\|_1 \to 0 \quad a.s.$$

Thus for P-almost all realisations $b^*(a,n,\cdot) \le f_n^* \le g^*(a,n,\cdot)$, one obtains that for all rational $a > 1$

$$\|g^*(a,n,\cdot) - a^{1+\gamma d} f\|_1 \to 0, \quad \|b^*(a,n,\cdot) - a^{-1-\gamma d} f\|_1 \to 0.$$

Let (n_k) be an arbitrary sequence of indices in $\mathbb{N}$. Then a subsequence (n_{k_l}) exists such that for all rational $a > 1$ (by Cantor's diagonal method)

$$g^*(a,n_{k_l},\cdot) \to a^{1+\gamma d} f, \quad b^*(a,n_{k_l},\cdot) \to a^{-1-\gamma d} f$$

λ-almost everywhere, thus $f^*_{n_{k_l}} \to f$ λ-almost everywhere and, by the Riesz-Vitali-Scheffé lemma, $\|f^*_{n_{k_l}} - f\|_1 \to 0$. Therefore $\|f_n^* - f\|_1 \to 0$, i.e., (30) is obtained. □

In order to establish in the situation of Theorem 8 strong consistency λ-almost everywhere for bounded K, in the α-mixing case one needs to strengthen the condition on γ to $0 < \gamma d < \min\{\frac{\alpha}{2}, \frac{\alpha}{\alpha+1}\}$, according to Irle (1997). Another result, where the freedom of choice in Theorem 8 is preserved, is given in the following corollary. The proof is similar to that of Theorem 8 and will be omitted. It uses variance inequalities of Peligrad (1992) and Rio (1993), respectively (see Lemma 3a,d).

Corollary 1. *Let the density K be as in Theorem 3. Assume further the conditions of Theorem 8 with (X_n) ρ-mixing or (X_n) weakly stationary and α-mixing with $\alpha_n = O(\delta^n)$ for some $\delta \in (0,1)$ and $\int K(x)^2 \log_+ K(x) \lambda(dx) < \infty$, further $0 < \gamma d < 1$. Then*

$$f_n(x) \to f(x) \quad a.s. \quad \text{mod } \lambda.$$

References

1. Chandra, T.K., Goswami, A. : Cesàro uniform integrability and the strong law of large numbers. *Sankhyā, Series A* **54**, 215-231 (1992).
2. Chandra, T.K., Goswami, A. : Corrigendum: Cesàro uniform integrability and the strong law of large numbers. *Sankhyā, Series A* **55**, 327-328 (1993).
3. Csörgő, S., Tandori, K., Totik, V. : On the strong law of large numbers for pairwise independent random variables. *Acta Mathematica Hungarica* **42**, 319-330 (1983).
4. Dehling, H., Philipp, W. : Almost sure invariance principles for weakly dependent vector-valued random variables. *Annals of Probability* **10**, 689-701 (1982).
5. Devroye, L. : On the almost everywhere convergence of nonparametric regression function estimates. *Annals of Statistics* **9**, 1310-1319 (1981).
6. Devroye, L., Györfi, L. : *Nonparametric Density Estimtion: The L_1-View.* Wiley, New York (1985).
7. Devroye, L., Györfi, L., Krzyżak, A., Lugosi, G. : On the strong universal consistency of nearest neighbor regression function estimates. *Annals of Statistics* **22**, 1371-1385 (1994).
8. Devroye, L., Krzyżak, A.: An equivalence theorem for L_1 convergence of the kernel regression estimate. *Journal of Statistical Planning and Inference* **23**, 71-82 (1989).
9. Devroye, L., Wagner, T.J.: Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Annals of Statistics* **8**, 231-239 (1980a).
10. Devroye, L., Wagner, T.J.: On the L_1-convergence of kernel estimators of regression functions with applications in discrimination. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **51**, 15-25 (1980b).
11. Doukhan, P.: *Mixing: Properties and Examples.* Springer, New York (1994).
12. Etemadi, N.: An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **55**, 119-122 (1981).
13. Etemadi, N.: On the law of large numbers for nonnegative random variables. *Journal of Multivariate Analysis* **13**, 187-193 (1983).

14. Ferrario, P.: *The Strong Law of Large Numbers and Applications*. Laurea Thesis, Politecnico di Milano (2004).
15. Frey, S.: *Konvergenzverhalten von Regressionsschätzungen unter Mischungsbedingungen.* Wissenschaftliche Arbeit, Universität Stuttgart (2007).
16. Gaposhkin, V.F.: Criteria for the strong law of large numbers for some classes of second-order stationary processes and homogeneous random fields. *Theory of Probability and its Applications* **22**, 286-310 (1977).
17. Greblicki, W., Krzyżak, A., Pawlak, M.; Distribution-free pointwise consistency of kernel regression estimate. *Annals of Statistics* **12**, 1570-1575 (1984).
18. Györfi, L.: Universal consistencies of a regression estimate for unbounded regression functions. *Nonparametric Functional Estimation and Related Topics* (Roussas, G., editor), 329-338. NATO ASI Series, Kluwer Academic Publishers, Dordrecht (1991).
19. Györfi, L., Härdle, W., Sarda, P., Vieu, P.: *Nonparametric Curve Extimation for Mixing Time Series.* Springer, New York (1989).
20. Györfi, L., Kohler, M., Krzyżak, A., Walk, H.: *A Distribution-Free Theory of Nonparametric Regression.* Springer Series in Statistics, Springer, New York (2002).
21. Györfi, L., Kohler, M., Walk, H.: Weak and strong universal consistency of semi-recursive partitioning and kernel regression estimates. *Statistics and Decisions* **16**, 1-18 (1998).
22. Györfi, L., Masry, E.: The L_1 and L_2 strong consistency of recursive kernel density estimation from dependent samples. *IEEE Transactions on Information Theory* **IT 36**, 531-539 (1990).
23. Hardly, G.H.: *Divergent Series.* Oxford University Press, London (1949).
24. Irle, A.: On consistency in nonparametric estimation under mixing conditions. *Journal of Multivariate Analysis* **60**, 123-147 (1997).
25. Karamata, J.: Einige Sätze über die Rieszschen Mittel. *Académie Royale Serbe, Bulletin de l'Académie des Sciences Mathématiques et Naturelles (Belgrade) A* **4**, 121-137 (1938).
26. Korevaar, J.: *Tauberian Theory: A Century of Developments.* Springer, Berlin (2004).
27. Kozek, A.S., Leslie, J.R., Schuster, E.F.:On a univeral strong law of large numbers for conditional expectations. *Bernoulli* **4**, 143-165 (1998).
28. Landau, E.: Über die Bedeutung einiger neuer Grenzwertsätze der Herren Hardy und Axer. *Prace matematyczno-fizyczne* **21**, 91-177 (1910).
29. Liebscher, E.: Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes and their Applications* **65**, 69-80 (1996).
30. Peligrad, M.: Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumptions. *Stochastics and Stochastics Reports* **40**, 147-168 (1992).
31. Rhomari, N.: Approximation et inégalités exponentielles pour les sommes de vecteurs aléatoires dépendants. *Comptes Rendus de l'Académie des Sciences Paris, Serie I* **334**, 149-154 (2002).
32. Rio, E.: Covariance inequalities for strongly mixing processes. *Annales de l'Institut Henri Poincaré, Section B* **29**, 587-597 (1993).
33. Schmidt, R.: Über divergente Folgen und lineare Mittelbildungen. *Mathematische Zeitschrift* **22**, 89-152 (1925).
34. Serfling, R.J.: Moment inequalities for the maximum cumulative sum. *Annals of Mathematical Statistics* **41**, 1227-1234 (1970a).
35. Serfling, R.J.: Convergence properties of S_n under moment restrictions. *Annals of Mathematical Statistics* **41**, 1235-1248 (1970b).
36. Spiegelman, C., Sacks, J.: Consistent window estimation in nonparametric regression. *Annals of Statistics* **8**, 240-246 (1980).
37. Steele, J.M.: An Efron-Stein inequality for nonsymmetric statistics. *Annals of Statistics* **14**, 753-758 (1986).
38. Stone, C.J.: Consistent nonparametric regression. *Annals of Statistics* **5**, 595-645 (1977).
39. Szász, O.: Verallgemeinerung und neuer Beweis einiger Sätze Tauberscher Art. *Sitzungsberichte der Bayerischen Akademie der Wissenschaften zu München, Mathematisch-Physikalische Klasse* **59**, 325-340 (1929).
40. Walk, H.: Strong universal pointwise consistency of recursive regression estimates. *Annals of the Institute of Statistical Mathematics* **53**, 691-707 (2001).
41. Walk, H.: Strong universal consistency of smooth kernel regression estimates. *Annals of the Institute of Statistical Mathematics* **57**, 665-685 (2005a).
42. Walk, H.: Strong laws of large numbers by elementary Tauberian arguments. *Monatshefte für Mathematik* **144**, 329-346 (2005b).
43. Walk, H.: Almost sure Cesàro and Euler summability of sequences of dependent random variables. *Archiv der Mathematik* **89**, 466-480 (2007).
44. Walk, H.: A universal strong law of large numbers for conditional expectations via nearest neighbors. *Journal of Multivariate Analysis* **99**, 1035-1050 (2008a).
45. Walk, H.: Strong consistency of kernel estimates under dependence. Submitted for publication (2008b).
46. Zeller, K., Beekmann, W.: *Theorie der Limitierungsverfahren,* 2. Auflage. Springer, Berlin (1970).

Harro Walk
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: walk@mathematik.uni-stuttgart.de
WWW: http://www.isa.uni-stuttgart.de/LstStoch/walk/

Erschienene Preprints ab Nummer 2007/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2009/003 *Walk, H.:* Strong laws of large numbers and nonparametric estimation

2009/002 *Leitner, F.:* The collapsing sphere product of Poincaré-Einstein spaces

2009/001 *Brehm, U.; Kühnel, W.:* Lattice triangulations of E^3 and of the 3-torus

2008/006 *Kohler, M.; Krzyżak, A.; Walk, H.:* Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps

2008/005 *Kaltenbacher, B.; Schöpfer, F.; Schuster, T.:* Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems

2008/004 *Leitner, F.:* Conformally closed Poincaré-Einstein metrics with intersecting scale singularities

2008/003 *Effenberger, F.; Kühnel, W.:* Hamiltonian submanifolds of regular polytope

2008/002 *Hertweck, M.; Hofert, C.R.; Kimmerle, W.:* Finite groups of units and their composition factors in the integral group rings of the groups $PSL(2,q)$

2008/001 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Two dimensional Berezin-Li-Yau inequalities with a correction term

2007/006 *Weidl, T.:* Improved Berezin-Li-Yau inequalities with a remainder term

2007/005 *Frank, R.L.; Loss, M.; Weidl, T.:* Polya's conjecture in the presence of a constant magnetic field

2007/004 *Ekholm, T.; Frank, R.L.; Kovarik, H.:* Eigenvalue estimates for Schrödinger operators on metric trees

2007/003 *Lesky, P.H.; Racke, R.:* Elastic and electro-magnetic waves in infinite waveguides

2007/002 *Teufel, E.:* Spherical transforms and Radon transforms in Moebius geometry

2007/001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions