The Annals of Probability
1991, Vol. 19, No. 2, 740–757

ON ITERATED LOGARITHM LAWS FOR LINEAR ARRAYS AND NONPARAMETRIC REGRESSION ESTIMATORS

BY PETER HALL

Australian National University

Laws of the iterated logarithm are derived for row sums of triangular arrays of independent random variables, in the context of nonparametric regression estimators. These laws provide exact strong convergence rates for kernel type nonparametric regression estimators. They apply to the important case where design points are conditioned upon, and permit the design to be multivariate. We impose minimal conditions on the error distribution; in fact, only finite variance is needed.

1. Introduction. The most common nonparametric regression model is

$$Y_i = g(x_i) + e_i, \qquad 1 \le i \le n,$$

where the pairs (x_i, Y_i) are observed, the x_i's are fixed d-variate vectors, the Y_i's are univariate random variables, the errors e_i are independent and identically distributed with zero mean and finite variance, and $g: \mathbb{R}^d \to \mathbb{R}$ is a smooth function which is to be estimated. The oldest type of estimator and one which is very commonly used in practice, is a kernel estimator

$$\hat{g}(x) = \sum_{i=1}^{n} Y_i L\{(x - x_i)/h_n\} \Big/ \sum_{i=1}^{n} L\{(x - x_i)/h_n\}. \tag{1.1}$$

Here $L: \mathbb{R}^d \to \mathbb{R}$ is a kernel function, typically a spherically symmetric d-variate probability density and h_n is the bandwidth or window size. Under mild regularity conditions, which include the assumption that $h_n \to 0$ and $nh_n^d \to \infty$, $\hat{g}(x)$ is consistent for $g(x)$ and has variance of size nh_n^d. The reader is referred to Härdle ([3], Section 3.1) and Prakasa Rao ([7], Section 4.2) for detailed accounts of these estimators and their properties.

Our aim in this paper is to give a precise description of the exact rate of strong convergence of $\hat{g}(x) - E\hat{g}(x)$, in the most common setting for regression problems: that where the design variables x_i are fixed, or conditioned upon, and represent a realization of a random sequence drawn from an unknown distribution. The convergence rate takes the form of a law of the iterated logarithm (LIL) for $\hat{g}(x) - E\hat{g}(x)$. Our result also applies to the case of regular design.

Härdle [2] has derived an LIL in the case where the design variables $x_i = X_i$ are univariate and are regarded as random. Härdle's proof is based on a bivariate version of the Komlós–Major–Tusnády approximation [4, 11], and is not adaptable to the important case where the x_i's are regarded as fixed.

Received April 1989; revised February 1990.
AMS 1980 *subject classifications*. Primary 60F15, 60G50; secondary 62G05.
Key words and phrases. Fixed design, law of the iterated logarithm, nonparametric regression, strong convergence rate, triangular array.

Stadtmüller [8] has used strong approximation techniques to prove an LIL for triangular arrays, which may in turn be employed to derive an LIL for an interpolation type of regression estimator having fixed, univariate design points. The contributions of Härdle and Stadtmüller are particularly noteworthy since the embedding techniques which they employ are new in the context of strong laws for weighted triangular arrays. However, neither method yields an LIL for the important case of kernel estimators with fixed design, and neither permits multivariate design. It seems very difficult to generalize the techniques of Härdle [2] and Stadtmüller [8] to this setting. In particular, multivariate design is excluded by Härdle's method because of limitations of the Komlós–Major–Tusnády embedding and cannot be treated by Stadtmüller's argument because of the difficulty of ordering multivariate design points. Therefore we have developed alternative techniques.

Our methods are closer to the classical techniques used to prove LIL's for triangular arrays than they are to the strong approximation approach. For LIL results about triangular arrays, see for example Gaposhkin [1] and Tomkins [9, 10]. However, none of these contributions or their precursors is applicable to the present case. The most general work has been done by Tomkins [10], but it covers only one part of the LIL, that where a lim sup is shown to be greater than or equal to (rather than equal to) a certain quantity. Furthermore, although the results and techniques of [10] might seem to be applicable to obtain at least this part of our LIL, they require regularity conditions which fail to hold in our case. For example, condition (a) on the growth of variance [10, page 308] does not hold. The work in [1, 9] is only distantly related to the regression case, being for a specific type of triangular array which is unrelated to the array arising in the regression setting. Moreover, the methods in [1, 9] demand at least the existence of a finite moment generating function, whereas our goal is to make only the minimum assumption of finite variance.

We have chosen the case of kernel estimators for the sake of definiteness and simplicity and because they are the estimator type most commonly used to introduce nonparametric regression. However our new techniques are readily adapted to other cases, including orthogonal series and histogram estimators. It seems to be quite difficult to formulate a general result which includes all these estimator types, in the context of fixed design.

Our main result is stated in Section 2. Section 3 presents preparatory lemmas needed in the proof, and the proof itself is given in Section 4.

2. Main theorem.

We assume that the kernel L is spherically symmetric and so is given by

$$L(x) = K(\|x\|), \tag{2.1}$$

where K is a univariate function and $\|x\|$ denotes the Euclidean norm of the d-vector x. In the case $d = 1$, we may take simply $K \equiv L$ and there the assumption of symmetry is not necessary for our method of proof. However, symmetric kernels are the overwhelmingly popular choice in applications. Moreover, the functions K which are used in practice are typically piecewise continuous; they are usually piecewise polynomials. We represent this property by asking that K be equal to a smooth function A_l on the interval $[a_{l-1}, a_l)$, for $1 \le l \le r$ say, where $0 = a_0 < \cdots < a_r < \infty$; and that K vanish on $[a_r, \infty)$. We also assume that for some constant $\chi > 1$, which may be chosen as large as necessary,

$$\sup_{0 \le y \le a_r} |A_l^{(j)}(y)| \le \chi^j, \quad \text{all } j \ge 0 \text{ and } 1 \le l \le r,$$

where $A_l^{(j)}$ denotes the jth derivative of the function A_l. Therefore, K may have any finite number of jump discontinuities within its support, but all derivatives of K must exist between those discontinuities.

For the sake of convenience we ask that L be rescaled so that $\int L = 1$. Of course, any scaling factor cancels from the definition (1.1) of the estimator $\hat{g}$.

We may and do assume without loss of generality that the scale of measurement has been chosen so that error variance is unity. Our only assumptions on the error distribution are the minimal ones,

$$E(e_i) = 0, \qquad E(e_i^2) = 1. \tag{2.2}$$

For the bandwidth h_n, we assume that

$$h_n \downarrow 0, \qquad nh_n^d \uparrow \infty, \qquad nh_n^d \asymp n^\alpha L(n), \tag{2.3}$$

where $0 < \alpha < 1$, L is slowly varying and $a_n \asymp b_n$ means that a_n/b_n and b_n/a_n are both bounded. The final condition in (2.3) is imposed only so that integral approximations may be made to series involving nh_n^d and may be relaxed.

In our main theorem we ask that the design points x_i come from a distribution which admits a density f in a neighbourhood of x, and that

$$f \text{ is continuous at } x \text{ and } f(x) > 0. \tag{2.4}$$

However, a remark following the theorem will point out that this method of generating the design may be generalized substantially without seriously hindering the proof.

We are now in a position to state our main theorem. Define $\hat{g}$ as at (1.1), with L given by (2.1). Assume the conditions on L stated in the first two paragraphs of this section and assume conditions (2.2), (2.3) and (2.4) on the error distribution, the bandwidth and the design distribution, respectively. We shall say that a result holds for a class of realizations of $x_1, x_2, \ldots$ having $\mathscr{X}$ probability 1 if that class has probability 1 in the distribution of random sequences $X_1, X_2, \ldots$ drawn from the design population.

THEOREM. *For a class of realizations $x_1, x_2, \ldots$ having $\mathscr{X}$ probability* 1,

$$\limsup_{n \to \infty} \left(nh_n^d / \log\log nh_n^d\right)^{1/2} \{\hat{g}(x) - E\hat{g}(x)\} = \left\{2\left(\int L^2\right)\Big/ f(x)\right\}^{1/2} \tag{2.5}$$

almost surely.

The theorem implies that

$$\liminf_{n \to \infty} \left(nh_n^d / \log\log nh_n^d\right)^{1/2} \{\hat{g}(x) - E\hat{g}(x)\} = -\left\{2\left(\int L^2\right)\Big/ f(x)\right\}^{1/2}$$

almost surely, as may be seen on replacing e_i by $-e_i$ in the regression model. Therefore,

$$\limsup_{n\to\infty} \left(nh_n^d/\log\log nh_n^d\right)^{1/2}|\hat{g}(x) - E\hat{g}(x)| = \left\{2\left(\int L^2\right)\Big/ f(x)\right\}^{1/2}$$

almost surely. It follows that the exact convergence rate of $\hat{g} - E\hat{g}$ is $O\{(nh_n^d/\log\log nh_n^d)^{1/2}\}$.

Note that our conditions on the kernel L permit L to be a high-order kernel [7, page 42]. However, we do ask that L be compactly supported. This assumption avoids the need for tail conditions on the design distribution and significantly reduces the length of proofs. Nevertheless, our techniques do extend to kernels with unbounded support. The trick is to approximate K by a compactly supported function and use techniques from the argument in part (b) of Step (i), Section 4, to control the difference between the estimator and its approximant. In particular, if we take K to be the standard normal density function and ask that, in addition to the other assumptions for the theorem, the parent distribution of the design variables satisfies $E(\|X\|^{\varepsilon}) < \infty$ for some $\varepsilon > 0$, then the theorem continues to hold.

Minor modifications to our proof show that the theorem remains true in the case where the design points x_i are chosen nonrandomly. For example, formula (2.5) remains true provided that $f(x)$ on the right-hand side is replaced by the value of

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\big(\text{number of points } x_i,\ 1\le i\le n, \text{ in a sphere of } d\text{-dimensional content } \varepsilon \text{ centred at } x\big)/(n\varepsilon). \tag{2.6}$$

Necessary regularity conditions are the same as for the theorem as stated, except that (2.4) should be replaced by the assumption that both the limits in (2.6) exist (finite) and the double limit is nonzero.

3. Preparatory lemmas.

Our proof of the main theorem is based on the four lemmas given here.

LEMMA 3.1. *Let $c_1,\dots,c_N$ be real numbers and $I_1,\dots,I_N$ be numbers taking only the values 0 and 1, such that $|c_i|I_i = |c_i|$ for $1 \le i \le N$. There exists an absolute constant A such that for all $u > 0$ and all $C_1, C_2 > 0$,*

$$P\left\{\left|\sum_{i=1}^N e_i c_i\right| > (C_1 + C_2)\left(\sum_{i=1}^N c_i^2\right)^{1/2} u\right\}$$
$$\le A\left[1 - \Phi(C_2 u) + C_2^{-3}\left(\sum_{i=1}^N c_i^2\right)^{-1/2}\right.$$
$$\left.\times\left(\sup_{1\le i\le N}|c_i|\right)u^{-3}E\{|e|^3 I(|e|\le\lambda)\} + \left(\sum_{i=1}^N I_i\right)P(|e|>\lambda)\right],$$

where $\lambda = (\sum I_i)^{1/2}/C_1$ *and* Φ *denotes the standard normal distribution function.*

PROOF. Put $e_i' = e_i I(|e_i| \le \lambda) - E\{e_i I(|e_i| \le \lambda)\}$. If $|e_i| \le \lambda$ for each i such that $I_i \ne 0$, then

$$\sum_{i=1}^{N} e_i c_i = \sum_{i=1}^{N} e_i' c_i + E\{eI(|e| \le \lambda)\} \sum_{i=1}^{N} c_i.$$

Now,

$$\left| E\{eI(|e| \le \lambda)\} \sum_{i=1}^{N} c_i \right| \le \lambda^{-1} \left(\sum_{i=1}^{N} c_i^2 \right)^{1/2} \left(\sum_{i=1}^{N} I_i \right)^{1/2} = C_1 \left(\sum_{i=1}^{N} c_i^2 \right)^{1/2}.$$

Hence if $u > 1$,

$$\begin{aligned} &P\left\{ \left| \sum_{i=1}^{N} e_i c_i \right| > (C_1 + C_2) \left(\sum_{i=1}^{N} c_i^2 \right)^{1/2} u \right\} \\ &\quad \le P\left\{ \left| \sum_{i=1}^{N} e_i' c_i \right| > C_2 \left(\sum_{i=1}^{N} c_i^2 \right)^{1/2} u \right\} + \left(\sum_{i=1}^{N} I_i \right) P(|e| > \lambda). \end{aligned} \tag{3.1}$$

Put $Y_i = e_i' c_i$ and $v = \operatorname{var}(\sum Y_i)$ and note that by the nonuniform version of Esseen's inequality ([6], point 23, page 132), there exists an absolute constant A_1 such that for every y,

$$\left| P\left(\left| \sum_{i=1}^{N} Y_i \right| > v^{1/2} y \right) - 2\{1 - \Phi(y)\} \right| \le A_1 (1 + |y|)^{-3} v^{-3/2} \sum_{i=1}^{N} E|Y_i|^3.$$

Take $w = C_2 (\sum c_i^2)^{1/2} u$ and $y = v^{-1/2} w$, obtaining

$$\left| P\left\{ \left| \sum_{i=1}^{N} e_i' c_i \right| > C_2 \left(\sum_{i=1}^{N} c_i^2 \right)^{1/2} u \right\} - 2\{1 - \Phi(v^{-1/2} w)\} \right| \le A_1 w^{-3} \sum_{i=1}^{N} E|Y_i|^3.$$

Now, $\sum E|Y_i|^3 \le E|e_1'|^3 \sum |c_i|^3$, $E|e_1'|^3 \le 8E\{|e|^3 I(|e| \le \lambda)\}$,

$$w^{-3} \sum_{i=1}^{N} |c_i|^3 \le C_2^{-3} u^{-3} \left(\sum_{i=1}^{N} c_i^2 \right)^{-1/2} \left(\sup_{1 \le i \le N} |c_i| \right)$$

and $v^{-1/2} w \ge C_2 u$. Therefore,

$$\begin{aligned} &P\left\{ \left| \sum_{i=1}^{N} e_i' c_i \right| > C_2 \left(\sum_{i=1}^{N} c_i^2 \right)^{1/2} u \right\} \\ &\quad \le 2\{1 - \Phi(C_2 u)\} + 8 A_1 C_2^{-3} u^{-3} \left(\sum_{i=1}^{N} c_i^2 \right)^{-1/2} \\ &\qquad \times \left(\sup_{1 \le i \le N} |c_i| \right) E\{|e|^3 I(|e| \le \lambda)\}. \end{aligned}$$

The lemma follows from this inequality and (3.1). □

LEMMA 3.2. *Let $Y_1, \dots, Y_N$ be independent random variables with zero means and finite variances, and put*

$$Z_i = \sum_{j=1}^{i} Y_j, \qquad z^2 = \operatorname{var}(Z_N).$$

Then if $u \ge \sqrt{2}$,

$$P\Big(\sup_{1 \le i \le N} |Z_i| > uz\Big) \le 2P\big\{|Z_N| > (u - \sqrt{2})z\big\}.$$

In particular,

$$P\Big(\sup_{1 \le i \le N} |Z_i| > 2uz\Big) \le 2P(|Z_N| > uz).$$

This is one version of an inequality due to Kolmogorov and it is proved in [5], page 260, for example.

LEMMA 3.3. *Let $Y_1, Y_2, \dots$ be independent random variables with zero means and finite variances and let $y_1, y_2, \dots$ be real numbers. For positive integers $m < m'$, put*

$$z^2 = \operatorname{var}\Bigg(\sum_{i=m+1}^{m'} Y_i\Bigg).$$

Then if $u \ge \sqrt{2}$,

$$P\Bigg\{\sup_{m<n<m'}\Bigg|\sum_{i=n+1}^{m'} Y_i I(y_i < h_n)\Bigg| > 2uz\Bigg\} \le 2P\Bigg(\Bigg|\sum_{i=m+1}^{m'} Y_i\Bigg| > uz\Bigg).$$

PROOF. Observe that

$$(y_i < h_n \text{ and } i > n) \Leftrightarrow (y_i < h_n \text{ and } h_i < h_n) \Leftrightarrow (z_i < h_n),$$

where $z_i = \max(y_i, h_i)$. Therefore,

$$\sum_{i=n+1}^{m'} Y_i I(y_i < h_n) = \sum_{i=m+1}^{m'} Y_i I(z_i < h_n).$$

Let $\{(Y_{(i)}, z_{(i)}),\ m < i \le m'\}$ denote the sequence $\{(Y_i, z_i),\ m < i \le m'\}$ ordered so that $z_{(m+1)} \le \cdots \le z_{(m')}$. Then

$$\begin{aligned}\sup_{m<n<m'}\Bigg|\sum_{i=n+1}^{m'} Y_i I(y_i < h_n)\Bigg| &= \sup_{m<n<m'}\Bigg|\sum_{i=m+1}^{m'} Y_{(i)} I(z_{(i)} < h_n)\Bigg| \\ &\le \sup_{m<n\le m'}\Bigg|\sum_{i=m+1}^{n} Y_{(i)}\Bigg|,\end{aligned}$$

and so

$$P\left[\sup_{m<n<m'}\left|\sum_{i=n+1}^{m'} Y_i I(Y_i < h_n)\right| > 2u\left\{\sum_{i=m+1}^{m'} E(Y_i^2)\right\}^{1/2}\right]$$

$$\le P\left[\sup_{m<n\le m'}\left|\sum_{i=m+1}^{n} Y_{(i)}\right| > 2u\left\{\sum_{i=m+1}^{m'} E(Y_{(i)}^2)\right\}^{1/2}\right]$$

$$\le 2P\left[\left|\sum_{i=m+1}^{m'} Y_{(i)}\right| > u\left\{\sum_{i=m+1}^{m'} E(Y_{(i)}^2)\right\}^{1/2}\right],$$

the last inequality following from Lemma 3.2. □

LEMMA 3.4. *For each $a > 0$, there exist constants $0 < C_1 < C_2 < \infty$ such that with $\mathscr{X}$ probability* 1,

$$C_1 nh_n^d \le \sum_{i=1}^{n} I(\|x - x_i\| < ah_n) \le C_2 nh_n^d$$

for all sufficiently large n.

PROOF. Let $X_1, X_2, \dots$ denote the independent and identically distributed random variables of which $x_1, x_2, \dots$ represents a realization. Since $f(x) > 0$, then

$$\delta_n = E\{I(\|x - X_1\| < ah_n)\} = P(\|x - X_1\| < ah_n) \sim C_3 h_n^d$$

as $n \to \infty$, where $C_3 > 0$. Therefore, the lemma will follow if we prove that with $Y_i = I(\|x - X_i\| < ah_n) - \delta_n$, we have

$$\left|\sum_{i=1}^{n} Y_i\right| = o\left\{\left(nh_n^d\right)^{1/2}\log\left(nh_n^d\right)\right\}, \tag{3.2}$$

with probability 1.

Put

$$b = \text{ess sup}|Y_i| \le 2,\ v_n = \text{var}\left(\sum Y_i\right) \le C_4 nh_n^d$$

and $u_n = \varepsilon(nh_n^d)^{1/2}\log(nh_n^d)$, where $\varepsilon > 0$. Then by Bernstein's inequality,

$$P\left(\left|\sum_{i=1}^{n} Y_i\right| > u_n\right) \le 2\exp\left[-(u_n^2/2)\{v_n + (bu_n)/3\}\right]$$

$$\le 2\exp\left[-C_5\{\log(nh_n^d)\}^2\right],$$

whence it follows that

$$\sum_{i=1}^{\infty} P\left(\left|\sum_{i=1}^{n} Y_i\right| > u_n\right) < \infty.$$

Result (3.2) now follows by the Borel–Cantelli lemma. □

4. Proof of main theorem. Observe that $\hat{g}(x) - E\hat{g}(x) = S_n/r_n$, where

$$S_n = \sum_{i=1}^{n} e_i K(\|x - x_i\|/h_n), \qquad r_n = \sum_{i=1}^{n} K(\|x - x_i\|/h_n);$$

and that with $\mathscr{X}$ probability 1, $r_n/(nh_n^d) \to f(x)$ as $n \to \infty$. The latter strong consistency result for density estimates is readily proved using techniques in, for example, [6], page 38.

Define

$$\sigma_n^2 = \operatorname{var}(S_n) = \sum_{i=1}^{n} K(\|x - x_i\|/h_n)^2.$$

Then it suffices to prove that for a class of realizations having $\mathscr{X}$ probability 1,

$$\sigma_n^2 \Big/ \left(nh_n^d \int K^2\right) \to f(x),$$

$$\limsup_{n\to\infty} S_n/\left(2\sigma_n^2 \log\log \sigma_n^2\right)^{1/2} = 1 \quad \text{almost surely.}$$

The first of these two results may once again be proved by methods which are standard for density estimators and so we establish only the second. This we do in two steps, which derive upper and lower bounds, respectively.

STEP (i). Upper bound. Write f.a.s.l.n as an abbreviation for for all sufficiently large n. In Step (i) we shall prove that for each $\varepsilon > 0$, with probability 1,

$$S_n/\left(2\sigma_n^2 \log\log \sigma_n^2\right)^{1/2} < 1 + \varepsilon \quad \text{f.a.s.l.}n. \tag{4.1}$$

Let $c > 1$ and write m_k for the integer part of c^k. Put

$$t_n = \left(2\sigma_n^2 \log\log \sigma_n^2\right)^{1/2}, \tag{4.2}$$

and note that t_n is increasing in n. Result (4.1) will follow if we prove that for each $\varepsilon > 0$ there exists $c > 1$, chosen sufficiently close to 1, such that

$$(4.3)\qquad S_{m_k}/t_{m_k} < 1 + \varepsilon \quad \text{f.a.s.l.}k,$$

$$(4.4)\qquad t_{m_k}^{-1} \sup_{m_k < n < m_{k+1}} |S_n - S_{m_k}| < \varepsilon \quad \text{f.a.s.l.}k.$$

Parts (a) and (b) below derive (4.3) and (4.4), respectively.

PART (a). Derivation of (4.3). In view of the Borel–Cantelli lemma, it suffices to prove that for each $\varepsilon > 0$,

$$(4.5)\qquad \sum_{k=1}^{\infty} P\left\{S_{m_k} > (1 + 2\varepsilon)t_{m_k}\right\} < \infty.$$

Assume that K vanishes outside $(-C_0, C_0)$. Bound the probability on the left-hand side of (4.5) by applying Lemma 3.1 with $c_i = K(\|x - x_i\|/h_{m_k})$, $I_i = I(\|x - x_i\|/h_{m_k} \le C_0)$, $u = u_k = (2\log\log\sigma_{m_k}^2)^{1/2}$, $C_1 = \varepsilon$ and $C_2 = 1 + \varepsilon$. Thus, in order to establish (4.5), it suffices to prove that for each $C_3 > 0$,

$$\sum_{k=1}^{\infty} \left[1 - \Phi\{(1 + \varepsilon)u_k\}\right] < \infty,$$

$$\sum_{k=1}^{\infty} \left(m_k h_{m_k}\right)^{-1/2} E\left[|e|^3 I\left\{|e| \le C_3\left(m_k h_{m_k}\right)^{1/2}\right\}\right] < \infty,$$

$$\sum_{k=1}^{\infty} \left(m_k h_{m_k}\right) P\left\{C_3|e| > \left(m_k h_{m_k}\right)^{1/2}\right\} < \infty.$$

(The $m_k h_{m_k}$ terms in the last two series arise because the quantity ΣI_i, which appears in the definition of λ in the Lemma 3.1, is here asymptotic to a constant multiple of $m_k h_{m_k}$.) This is readily accomplished by making integral approximations to the series.

PART (b). Derivation of (4.4). Put $m = m_k$, $m' = m_{k+1}$. If $m < n < m'$, then

$$S_n - S_m = \sum_{i=1}^{m} e_i\{K(\|x - x_i\|/h_n) - K(\|x - x_i\|/h_m)\}$$
$$+ \sum_{i=m+1}^{n} e_i K(\|x - x_i\|/h_n).$$

Our assumption that $K(y) = \Sigma A_l(y) I(a_{l-1} \le y < a_l)$ for smooth functions A_l permits us to use the representation $K(y) = \Sigma B_l(y) I(y < a_l)$, where B_l is a linear combination of the A_l's and therefore satisfies the conditions imposed on the A_l's. Hence to prove (4.4), it suffices to show that for $c > 1$ sufficiently close to 1,

$$t_{m_k}^{-1} \sup_{m_k<n<m_{k+1}} \left| \sum_{i=1}^{m_k} e_i \{ B(\|x-x_i\|/h_n) I(\|x-x_i\|/h_n < a) - B(\|x-x_i\|/h_{m_k}) I(\|x-x_i\|/h_{m_k} < a) \} \right| \le \varepsilon \quad \text{f.a.s.l.}k,$$

$$t_{m_k}^{-1} \sup_{m_k<n<m_{k+1}} \left| \sum_{i=m_k+1}^{n} e_i B(\|x-x_i\|/h_n) I(\|x-x_i\|/h_n < a) \right| \le \varepsilon \quad \text{f.a.s.l.}k,$$

where (B, a) denotes a generic (B_l, a_l). These two results (for 2ε rather than ε) will follow if we prove that

$$t_{m_k}^{-1} \sup_{m_k<n<m_{k+1}} \left| \sum_{i=1}^{m_k} e_i \{ B(\|x-x_i\|/h_n) - B(\|x-x_i\|/h_{m_k}) \} \times I(\|x-x_i\|/h_n < a) \right| \le \varepsilon \quad \text{f.a.s.l.}k, \tag{4.6}$$

$$t_{m_k}^{-1} \sup_{m_k<n<m_{k+1}} \left| \sum_{i=1}^{m_k} e_i B(\|x-x_i\|/h_{m_k}) \times I(\|x-x_i\|/h_{m_k} \le a < \|x-x_i\|/h_n) \right| \le \varepsilon \quad \text{f.a.s.l.}k,$$

$$t_{m_k}^{-1} \sup_{m_k<n<m_{k+1}} \left| \sum_{i=n+1}^{m_{k+1}} e_i \{ B(\|x-x_i\|/h_{m_{k+1}}) - B(\|x-x_i\|/h_n) \} \times I(\|x-x_i\|/h_n < a) \right| \le \varepsilon \quad \text{f.a.s.l.}k, \tag{4.7}$$

$$t_{m_k}^{-1} \sum_{m_k<n<m_{k+1}} \left| \sum_{i=n+1}^{m_{k+1}} e_i B(\|x-x_i\|/h_{m_{k+1}}) I(\|x-x_i\|/h_n < a) \right| \le \varepsilon \quad \text{f.a.s.l.}k.$$

If

$$(h_{m_k} - h_{m_{k+1}})/h_{m_{k+1}} < \theta, \qquad k \ge k_0, \tag{4.8}$$

then for $k \ge k_0$, the left-hand sides of (4.6) and (4.7) are dominated by

$$t_{m_k}^{-1} \sum_{j=1}^{\infty} (\theta^j/j!) \sup_{m_k<n<m_{k+1}} \left| \sum_{i=1}^{m_k} e_i B^{(j)}(\|x-x_i\|/h_{m_k}) I(\|x-x_i\|/h_n < a) \right|,$$

$$t_{m_k}^{-1} \sum_{j=1}^{\infty} (\theta^j/j!) \sup_{m_k<n<m_{k+1}} \left| \sum_{i=n+1}^{m_{k+1}} e_i B^{(j)}(\|x-x_i\|/h_{m_{k+1}}) I(\|x-x_i\|/h_n < a) \right|,$$

respectively. Therefore, result (4.4) will follow if we prove that for a sequence of positive constants $\{\tau_{jk};\ j \geq 1,\ k \geq 1\}$ satisfying

$$\sum_{j=1}^{\infty} (\theta^j/j!)\tau_{jk} \leq \varepsilon, \tag{4.9}$$

we have

$$\begin{aligned}\sum_{j=1}^{\infty}\sum_{k=k_0}^{\infty} P\Bigg\{ \sup_{m_k<n<m_{k+1}} \Bigg| \sum_{i=1}^{m_k} e_i B^{(j)}\big(\|x-x_i\|/h_{m_k}\big) \\ \times I\big(\|x-x_i\|/h_n < a\big)\Bigg| > t_{m_k}\tau_{jk}\Bigg\} < \infty,\end{aligned} \tag{4.10}$$

$$\begin{aligned}\sum_{k=k_0}^{\infty} P\Bigg\{ \sup_{m_k<n<m_{k+1}} \Bigg| \sum_{i=1}^{m_k} e_i B\big(\|x-x_i\|/h_{m_k}\big) \\ \times I\big(\|x-x_i\|/h_{m_k} \leq a < \|x-x_i\|/h_n\big)\Bigg| > t_{m_k}\varepsilon\Bigg\} < \infty,\end{aligned} \tag{4.11}$$

$$\begin{aligned}\sum_{j=0}^{\infty}\sum_{k=k_0}^{\infty} P\Bigg\{ \sup_{m_k<n<m_{k+1}} \Bigg| \sum_{i=n+1}^{m_{k+1}} e_i B^{(j)}\big(\|x-x_i\|/h_{m_{k+1}}\big) \\ \times I\big(\|x-x_i\|/h_n < a\big)\Bigg| > t_{m_k}\tau_{jk}\Bigg\} < \infty,\end{aligned} \tag{4.12}$$

where in (4.12), we take $\tau_{0k} = \varepsilon$.

The next step is to remove the suprema from the left-hand sides of (4.10)–(4.12). In the case of (4.10) and (4.11), this may be done by using Lemma 3.2, which applies to suprema of sequences of consecutive partial sums. Each of the suprema in (4.10) and (4.11) may be put in that form by reordering the indices i, $1 \leq i \leq m_k$, so that the differences $\|x - x_i\|$ are monotone. Arguing thus, and defining

$$w_{jk}^2 = \sum_{i=1}^{m_k} B^{(j)}\big(\|x-x_i\|/h_{m_k}\big)^2 I\big(\|x-x_i\|/h_{m_k} < a\big),$$

$$w_k^2 = \sum_{i=1}^{m_k} B\big(\|x-x_i\|/h_{m_k}\big)^2 I\big(\|x-x_i\|/h_{m_k} \leq a < \|x-x_i\|/h_{m_{k+1}}\big),$$

we see from Lemma 3.2 that if

$$t_{m_k}\tau_{jk} > 2\sqrt{2}\, w_{jk}, \qquad t_{m_k}\varepsilon > 2\sqrt{2}\, w_k, \tag{4.13}$$

then (4.10) and (4.11) are implied by

$$\sum_{j=1}^{\infty}\sum_{k=k_0}^{\infty} P\left\{\left|\sum_{i=1}^{m_k} e_i B^{(j)}(\|x-x_i\|/h_{m_k}) \times I(\|x-x_i\|/h_{m_k}<a)\right| > t_{m_k}\tau_{jk}/2\right\} < \infty, \tag{4.14}$$

$$\sum_{k=k_0}^{\infty} P\left\{\left|\sum_{i=1}^{m_k} e_i B(\|x-x_i\|/h_{m_k}) \times I(\|x-x_i\|/h_{m_k} \le a < \|x-x_i\|/h_{m_{k+1}})\right| > t_{m_k}\varepsilon/2\right\} < \infty. \tag{4.15}$$

The supremum in (4.12) cannot be rendered into a supremum of consecutive partial sums simply by changing the order of the indices i. In this case we resort to Lemma 3.3 rather than Lemma 3.2. Arguing thus, and defining

$$v_{jk}^2 = \sum_{i=m_k+1}^{m_{k+1}} B^{(j)}(\|x-x_i\|/h_{m_{k+1}})^2 I(\|x-x_i\|/h_{m_k}<a),$$

we see that if $t_{m_k}\tau_{jk} > 2\sqrt{2}\,v_{jk}$, then (4.12) is implied by

$$\sum_{j=0}^{\infty}\sum_{k=k_0}^{\infty} P\left\{\left|\sum_{i=m_k+1}^{m_{k+1}} e_i B^{(j)}(\|x-x_i\|/h_{m_{k+1}}) \times I(\|x-x_i\|/h_{m_k}<a)\right| > t_{m_k}\tau_{jk}/2\right\} < \infty. \tag{4.16}$$

Result (4.4) will follow if we prove (4.14)–(4.16).

We establish (4.14)–(4.16) by using Lemma 3.1 to bound the respective probabilities. The method is similar in each case, and so for the sake of brevity we shall derive only (4.14). In Lemma 3.1, take $\chi > 1$ and put $N = N_k = m_k$,

$$c_i = c_{ijk} = B^{(j)}(\|x-x_i\|/h_{m_k}) I(\|x-x_i\|/h_{m_k}<a),$$

$$I_i = I_{ik} = I(\|x-x_i\|/h_{m_k}<a),$$

$$C_1 = C_{1j} = \chi^{-j}, \qquad C_2 = C_{2j} = \chi^{j},$$

$$u = u_{jk} = \left(\sum_{i=1}^{N} c_i^2\right)^{-1/2}\left(\sum_{i=1}^{N} I_i\right)^{1/2}\left(\sup_{1\le i\le N}|c_i|\right)\chi^j\left(\log\log m_k h_{m_k}^d\right)^{1/2}.$$

Furthermore, define

$$\tau_{jk} = 2t_{m_k}^{-1}(\chi^{-j}+\chi^{j})\left(\sum_i I_{ik}\right)^{1/2}\left(\sup_i|c_{ijk}|\right)\chi^j\left(\log\log m_k h_{m_k}^d\right)^{1/2}. \tag{4.17}$$

Then

$$t_{m_k}\tau_{jk}/2 = (C_1 + C_2)\left(\sum_i c_i^2\right)^{1/2} u_{jk},$$

and so by Lemma 3.1, the left-hand side of (4.14) is dominated by

$$(4.18)\qquad \begin{aligned} A\sum_{j=1}^{\infty}\sum_{k=k_0}^{\infty}\Bigg(&1 - \Phi\left(\chi^j u_{jk}\right) \\ &+\chi^{-3j}\left(\sum_i I_{ik}\right)^{-1/2} E\left[|e|^3 I\left\{|e| \le \chi^j\left(\sum_i I_{ik}\right)^{1/2}\right\}\right] \\ &+\left(\sum_i I_{ik}\right)P\left\{|e| > \chi^j\left(\sum_i I_{ik}\right)^{1/2}\right\}\Bigg), \end{aligned}$$

provided that (4.9) and (4.13) hold with the definition (4.17) of τ_{jk}. In obtaining the quantity at (4.18), we have used the following argument to bound the coefficient of the third term:

$$\begin{aligned} &\left(\sum_i c_i^2\right)^{-1/2}\left(\sup_{1\le i\le N}|c_i|\right)u^{-3} \\ &= \left(\sum_i c_i^2\right)\left(\sup_{1\le i\le N}|c_i|\right)^{-2}\left(\sum_i I_i\right)^{-3/2}\chi^{-3j}\left(\log\log m_k h_{m_k}^d\right)^{-3/2} \\ &\le \left(\sum_i c_i^2\right)\left(\sup_{1\le i\le N}|c_i|\right)^{-2}\left(\sum_i I_i\right)^{-3/2} \\ &\le \left(\sum_i I_i\right)^{-1/2}. \end{aligned}$$

In view of Lemma 3.4, there exist positive constants $0 < C_3 < C_4 < \infty$ such that

$$(4.19)\qquad C_3 m_k h_{m_k}^d \le \sum_i I_{ik} \le C_4 m_k h_{m_k}^d$$

for $k \ge k_0$. If χ is chosen so large that

$$\sup_{0\le y\le a_r}|B^{(j)}(y)| \le \chi^j,$$

then the right-hand side of (4.17) is dominated by

$$4C_4^{1/2}\chi^{3j}t_{m_k}^{-1}\left(m_k h_{m_k}^d \log\log m_k h_{m_k}^d\right)^{1/2} \le C_5\chi^{3j}.$$

If we choose $c > 1$ so close to 1 and k_0 so large that (4.8) holds with

$$\theta = \min\left[\chi^{-6}, \left\{\log\left(1 + C_5^{-1}\varepsilon\right)\right\}^2\right],$$

then we shall have

$$\sum_{j=1}^{\infty} (\theta^j/j!)\tau_{jk} \le C_5 \sum_{j=1}^{\infty} \theta^{j/2}/j! = C_5\{\exp(\theta^{1/2}) - 1\} \le \varepsilon,$$

which ensures (4.9) with τ_{jk} given by (4.17). Similarly it may be proved that (4.13) holds for this τ_{jk}, provided χ is chosen sufficiently large. Therefore it suffices to prove that for large χ, the series in (4.18) converges. This may be done by integral approximations to the three series which comprise (4.18), using (4.19). To treat the first of the three series, note that $u_{jk} \ge (\log\log m_k h_{m_k}^d)^{1/2}$.

STEP (ii). Lower bound. Write i.o. for infinitely often. In Step (ii), we shall prove that for each $0 < \varepsilon < 1$, with probability 1,

$$S_n/\left(2\sigma_n^2 \log\log \sigma_n^2\right)^{1/2} > 1 - \varepsilon \quad \text{i.o.} \tag{4.20}$$

Write m_k for the integer part of c^k and define t_n as at (4.2). Result (4.20) will follow if we prove that for each $\varepsilon > 0$, there exists $c > 1$, chosen sufficiently large, such that

$$|S_{m_k}|/t_{m_{k+1}} < \varepsilon \quad \text{f.a.s.l.}k, \tag{4.21}$$

$$(S_{m_{k+1}} - S_{m_k})/t_{m_{k+1}} > 1 - \varepsilon \quad \text{i.o.} \tag{4.22}$$

The proof of (4.21) is very similar to our derivation of (4.3) in Part (a) of Step (i), and so is not given here. We prove only (4.22).

Observe that

$$S_{m_{k+1}} - S_{m_k} = \sum_{i=1}^{m_k} e_i\left\{K\left(\|x - x_i\|/h_{m_{k+1}}\right) - K\left(\|x - x_i\|/h_{m_k}\right)\right\} + \sum_{i=m_k+1}^{m_{k+1}} e_i K\left(\|x - x_i\|/h_{m_{k+1}}\right).$$

Therefore (4.22) will follow if we prove that for each $\varepsilon > 0$, we have for $c > 1$ sufficiently large,

$$t_{m_{k+1}}^{-1}\left|\sum_{i=1}^{m_k} e_i\left\{K\left(\|x - x_i\|/h_{m_{k+1}}\right) - K\left(\|x - x_i\|/h_{m_k}\right)\right\}\right| < \varepsilon \quad \text{f.a.s.l.}k, \tag{4.23}$$

$$t_{m_{k+1}}^{-1} \sum_{i=m_k+1}^{m_{k+1}} e_i I\left(\|x - x_i\|/h_{m_{k+1}}\right) > 1 - \varepsilon \quad \text{i.o.} \tag{4.24}$$

Result (4.23) is a consequence of

$$\sum_{k=1}^{\infty} P\left[\left|\sum_{i=1}^{m_k} e_i\left\{K\left(\|x - x_i\|/h_{m_{k+1}}\right) - K\left(\|x - x_i\|/h_{m_k}\right)\right\}\right| > \varepsilon t_{m_{k+1}}\right] < \infty,$$

which follows from Lemma 3.1. The method of proof is similar to that employed earlier to derive (4.14). Hence we confine attention to proving (4.24).

For distinct values of k, the series on the left-hand side of (4.24) is stochastically independent. Therefore (4.24) will follow via the Borel–Cantelli lemma if we prove that for $c > 1$ sufficiently large,

$$\sum_{k=1}^{\infty} P\{T_k > (1-\varepsilon)t_{m_{k+1}}\} = \infty, \tag{4.25}$$

where

$$T_k = \sum_{i=m_k+1}^{m_{k+1}} e_i K(\|x - x_i\|/h_{m_{k+1}}).$$

The remainder of our proof is dedicated to deriving (4.25). First we need an estimate of the distribution function of T_k. Put $\lambda_k = (m_{k+1} h^d_{m_{k+1}})^{1/2}$,

$$e_{ik} = e_i I(|e_i| \le \lambda_k) - E\{e_i I(|e_i| \le \lambda_k)\},$$

$$T'_k = \sum_{i=m_k+1}^{m_{k+1}} e_{ik} K(\|x - x_i\|/h_{m_{k+1}}),$$

$$\mu_k = E\{eI(|e| \le \lambda_k)\} \sum_{i=m_k+1}^{m_{k+1}} K(\|x - x_i\|/h_{m_{k+1}}),$$

$$v_k = \operatorname{var}(e_{1k}) \sum_{i=m_k+1}^{m_{k+1}} K(\|x - x_i\|/h_{m_{k+1}})^2.$$

Assume K vanishes outside $(-C_0, C_0)$. Then

$$\begin{aligned} &\sup_{-\infty<y<\infty} |P(T_k \le v_k^{1/2} y + \mu_k) - P(T'_k \le v_k^{1/2} y)| \\ &\qquad \le \left\{ \sum_{i=m_k+1}^{m_{k+1}} I(\|x - x_i\|/h_{m_{k+1}} \le C_0) \right\} P(|e| > \lambda_k). \end{aligned} \tag{4.26}$$

By the Berry–Esseen theorem ([6], page 111), there exists an absolute constant $A > 0$ such that

$$\begin{aligned} &\sup_{-\infty<y<\infty} |P(T'_k \le v_k^{1/2} y) - \Phi(y)| \\ &\qquad \le A v_k^{-3/2} \sum_{i=m_k+1}^{m_{k+1}} E\{|e_{ik} K(\|x - x_i\|/h_{m_{k+1}})|^3\} \\ &\qquad \le 8A(\sup|K|^3) v_k^{-3/2} \sum_{i=m_k+1}^{m_{k+1}} I(\|x - x_i\|/h_{m_{k+1}} \le C_0). \end{aligned} \tag{4.27}$$

Next we estimate the right-hand sides of (4.26) and (4.27). The argument employed to prove Lemma 3.4 may be used to show that with $\mathscr{X}$ probability 1, as $k \to \infty$,

$$\sum_{i=m_k+1}^{m_{k+1}} I\left(\|x - x_i\|/h_{m_{k+1}} \le C_0\right) \sim C_1(m_{k+1} - m_k)h_{m_{k+1}}^d \sim C_1(1 - c^{-1})m_{k+1}h_{m_{k+1}}^d, \tag{4.28}$$

$$\sum_{i=m_k+1}^{m_{k+1}} |K(\|x - x_i\|/h_{m_{k+1}})| \le (\sup|K|) \sum_{i=m_k+1}^{m_{k+1}} I\left(\|x - x_i\|/h_{m_{k+1}} \le C_0\right) \sim C_1(1 - c^{-1})m_{k+1}h_{m_{k+1}}^d, \tag{4.29}$$

$$\sum_{i=m_k+1}^{m_{k+1}} K\left(\|x - x_i\|/h_{m_{k+1}}\right)^2 \sim C_2(m_{k+1} - m_k)h_{m_{k+1}}^d \sim C_2(1 - c^{-1})m_{k+1}h_{m_{k+1}}^d, \tag{4.30}$$

$$\sigma_{m_{k+1}}^2 = \sum_{i=1}^{m_{k+1}} K\left(\|x - x_i\|/h_{m_{k+1}}\right)^2 \sim C_2 m_{k+1}h_{m_{k+1}}^d,$$

$$t_{m_{k+1}} = \left(2\sigma_{m_{k+1}}^2 \log\log \sigma_{m_{k+1}}^2\right)^{1/2} \sim \left\{2C_2 m_{k+1}h_{m_{k+1}}^d \log\log\left(m_{k+1}h_{m_{k+1}}^d\right)\right\}^{1/2}, \tag{4.31}$$

where $C_1, C_2 > 0$ do not depend on c. From (4.29) and (4.31), it follows that

$$|\mu_k| \le \lambda_k^{-1}E\{e^2 I(|e| > \lambda_k)\} \sum_{i=m_k+1}^{m_{k+1}} |K(\|x - x_i\|/h_{m_{k+1}})| = o(t_{m_{k+1}}); \tag{4.32}$$

from (4.30),

$$v_k \sim \sum_{i=m_k+1}^{m_{k+1}} K\left(\|x - x_i\|/h_{m_{k+1}}\right)^2 \sim C_2(1 - c^{-1})m_{k+1}h_{m_{k+1}}^d; \tag{4.33}$$

from (4.26) and (4.28),

$$\sup_{-\infty < y < \infty} |P(T_k \le v_k^{1/2}y + \mu_k) - P(T_k' \le v_k^{1/2}y)| \le C_3\left(m_{k+1}h_{m_{k+1}}^d\right) P\left\{|e| > \left(m_{k+1}h_{m_{k+1}}^d\right)^{1/2}\right\}; \tag{4.34}$$

from (4.27), (4.28) and (4.33),

$$\sup_{-\infty < y < \infty} |P(T_k' \le v_k^{1/2}y) - \Phi(y)| \le C_3\left(m_{k+1}h_{m_{k+1}}^d\right)^{-1/2} E\left[|e|^3 I\left\{|e| \le \left(m_{k+1}h_{m_{k+1}}^d\right)^{1/2}\right\}\right], \tag{4.35}$$

where in (4.34) and (4.35), C_3 depends on c; from (4.34) and (4.35),

$$\begin{aligned} &\sup_{-\infty<y<\infty} |P(T_k > v_k^{1/2}y + \mu_k) - \{1 - \Phi(y)\}| \\ (4.36)\quad &\le C_3\Big(\big(m_{k+1}h^d_{m_{k+1}}\big)P\Big\{|e| > \big(m_{k+1}h^d_{m_{k+1}}\big)^{1/2}\Big\} \\ &\qquad + \big(m_{k+1}h^d_{m_{k+1}}\big)^{-1/2}E\Big[|e|^3 I\Big\{|e| \le \big(m_{k+1}h^d_{m_{k+1}}\big)^{1/2}\Big\}\Big]\Big) \\ &\equiv \tau_k, \end{aligned}$$

say; and from (4.31)–(4.33),

$$\begin{aligned} (4.37)\quad y_k &\equiv \big\{(1-\varepsilon)t_{m_{k+1}} - \mu_k\big\}/v_k^{1/2} \sim (1-\varepsilon)t_{m_{k+1}}/v_k^{1/2} \\ &\sim (1-\varepsilon)(1-c^{-1})^{-1/2}\Big\{2\log\log\big(m_{k+1}h^d_{m_{k+1}}\big)\Big\}^{1/2}. \end{aligned}$$

Take $y = y_k$ in (4.36) and choose $c > 1$ so large that $(1-\varepsilon)(1-c^{-1})^{-1/2} \le 1 - (2\varepsilon/3)$. Then for large k, by (4.37),

$$y_k \le \{1 - (\varepsilon/3)\}\Big\{2\log\log\big(m_{k+1}h^d_{m_{k+1}}\big)\Big\}^{1/2} \equiv y_k',$$

and so by (4.36),

$$(4.38)\qquad P\big\{T_k > (1-\varepsilon)t_{m_{k+1}}\big\} \ge 1 - \Phi(y_k') - \tau_k.$$

It may be shown by integral approximations to two series that $\Sigma\tau_k < \infty$. The relation

$$1 - \Phi(y) \sim y^{-1}(2\pi)^{-1/2}\exp(-y^2/2),$$

valid as $y \to \infty$, allows that $\Sigma\{1 - \Phi(y_k')\} = \infty$. The desired result (4.25) follows from these results and (4.38). □

Acknowledgment. I am grateful to the referee for his helpful comments.

REFERENCES

[1] GAPOSHKIN, V. F. (1965). The law of the iterated logarithm for Cesaro's and Abel's methods of summation. *Theory Probab. Appl.* **10** 411–420.

[2] HÄRDLE, W. (1984). A law of the iterated logarithm for nonparametric regression function estimators. *Ann. Statist.* **12** 624–635.

[3] HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.

[4] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent rv's and the sample df. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

[5] LOÈVE, M. (1977). *Probability Theory* **1**, 4th ed. Springer, Berlin.

[6] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

[7] PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic, New York.

[8] STADTMÜLLER, U. (1984). A note on the law of the iterated logarithm for weighted sums of random variables. *Ann. Probab.* **12** 35–44.

[9] TOMKINS, R. J. (1971). An iterated logarithm theorem for some weighted averages of independent random variables. *Ann. Math. Statist.* **42** 760–763.

[10] TOMKINS, R. J. (1974). On the law of the iterated logarithm for double sequences of random variables. *Z. Wahrsch. Verw. Gebiete* **30** 303–314.

[11] TUSNÁDY, G. (1977). A remark on the approximation of the sample df in the multivariate case. *Period. Math. Hungar.* **8** 53–55.

DEPARTMENT OF STATISTICS
THE AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA