The Annals of Statistics
1984, Vol. 12, No. 2, 624–635

A LAW OF THE ITERATED LOGARITHM FOR NONPARAMETRIC REGRESSION FUNCTION ESTIMATORS[1]

BY WOLFGANG HÄRDLE

Universität Heidelberg

We study the estimation of a regression function by two classes of estimators, the Nadaraya-Watson Kernel type estimators and the orthogonal polynomial estimators. We obtain sharp pointwise rates of strong consistency by establishing laws of the iterated logarithm for the two classes of estimators. These results parallel those of Hall (1981) on density estimation and extend those of Noda (1976) on strong consistency of kernel regression estimators.

1. Introduction and background. Let (X, Y), (X_i, Y_i), $i = 1, 2, \cdots$ be i.i.d. bivariate random variables with common joint distribution $F(x, y)$ and joint density $f(x, y)$. Let $f_X(x)$ be the marginal density of X and let $m(x) = E(Y \mid X = x) = \int yf(x, y)\, dy/f_X(x)$ be the regression of Y on X. In the present paper we obtain sharp pointwise rates of strong consistency for the following type of regression estimator

$$m_n(x) = n^{-1} \sum_{i=1}^{n} K_{r(n)}(x; X_i) Y_i \tag{1.1}$$

where $\{K_r\colon r \in I\}$ denotes a sequence of "delta functions" (or kernel sequence).

Many nonparametric estimators of $m(x)$ have this form, for instance, the Nadaraya-Watson kernel estimator (more generally estimators based on delta function sequences, as introduced by Watson and Leadbetter, 1964) or orthogonal polynomial estimators.

Nadaraya (1964) and Watson (1964) independently introduced a kernel type variant of (1.1) and demonstrated weak pointwise consistency. Rosenblatt (1969) obtained the bias, variance and asymptotic distribution of kernel type regression estimators. Schuster (1972) and Johnston (1979) demonstrated the multivariate normality at a finite number of distinct points. The strong pointwise consistency (without rates) of the Nadaraya-Watson estimator was shown by Noda (1976). For this particular kernel type estimator Collomb (1979) gave necessary and sufficient conditions on the sequence $\{K_{r(n)}\}$ for strong consistency of m_n. Stone (1977) gave general conditions on the weights $K_r(x; X_i)$ for $m_n(x)$ to be consistent in L^r, i.e. for $E \mid m_n(X) - m(X) \mid^r \to 0$. From his conditions, however, it is not clear when the Nadaraya-Watson kernel sequence is consistent (Stone, 1977, page 607).

Recently, Schuster and Yakowitz (1979) derived uniform consistency on a finite interval for a kernel type estimator. Wandl (1980) and Johnston (1982) studied the global deviation and Revesz (1979) obtained analogous results includ-

Received February 1983; revised November 1983.

[1] Research supported by the Deutsche Forschungsgemeinschaft, SFB 123 and by the Scientific Research Contract AFOSR-F49620-82-C-0009.

AMS 1980 *subject classification.* Primary 60F10; secondary 60G15, 62G05.

Key words and phrases. Nonparametric regression function estimation, law of the iterated logarithm, kernel estimation, orthogonal polynomial estimation.

ing nearest neighbor regression estimators. In addition, Wandl (1980) obtained rates of uniform consistency, but under the rather restrictive assumption that the marginal distribution of Y has bounded support. The assumptions in Mack and Silverman (1982), who show weak and strong uniform consistency on a bounded interval of the Nadaraya-Watson kernel estimator, are less restrictive than in Wandl (1980); the difficulties with an unbounded support of Y are overcome by a truncation argument. A similar technique, together with strong approximations of the two dimensional empirical process, will be used in the present paper. Different criteria measuring the closeness of m_n to m, including the L_1-distance, for kernel type estimators were considered by Devroye (1978, 1981) and by Devroye and Wagner (1980a, b).

The method of orthogonal polynomial estimation was originally introduced by Čencov (1962) for density estimation. Rutkowski (1982a, b) defined a regression estimator based on orthogonal polynomials in the case of fixed design variables X. He also presented conditions for (weak) consistency and discussed the applications of such estimators to a broad class of system identification problems. For more work and related problems concerning both kernel type and orthogonal polynomial type estimators, we refer to the review article of Collomb (1981).

In the present paper we show a law of the iterated logarithm for the centered estimate

$$m_n(x) - Em_n(x). \tag{1.2}$$

This result thus gives the exact order of convergence of $m_n(x) - Em_n(x)$. For statistical interpretations it is desirable to have exact pointwise strong convergence rates for $m_n(x) - m(x)$, but since the bias is purely analytically handled, it suffices to consider (1.2). The handling with the bias terms using different smoothness assumptions on m and K_r is delayed to the sections where we apply the general result of Section 2. In Section 4 we show a law of the iterated logarithm for the Nadaraya-Watson kernel type estimator and for a related estimator that is useful if we know the marginal density f_X of X. In Section 5 we derive an analogous result for estimators based on orthogonal polynomials.

As a footnote, we would like to mention some related works on density estimation. These include among others Wegman and Davies (1979), Hall (1981), Stute (1982).

2. A law of the iterated logarithm for a special triangular array. Let $(X_1, Y_1), (X_2, Y_2), \cdots$ be a sequence of independent and identically distributed random variables with probability density function $f(x, y)$ and cumulative distribution function $F(x, y)$ and $EY^2 < \infty$. As in (1.1), let $\{K_r: r \in I\}$ be a sequence of real valued functions each of bounded variation and define

$$S_n(r) = \sum_{i=1}^{n} \{K_r(X_i)Y_i - E[K_r(X_i)Y_i]\}.$$

Note that this sum is a multiple of (1.2), and that we omitted the dependence on the design point x for convenience. Define also

$$\sigma(r, s) = \operatorname{cov}\{K_r(X)Y, K_s(X)Y\} \quad \text{and} \quad \sigma^2(r) = \sigma(r, r).$$

We will now establish conditions under which $S_n(r) = n[m_n(x) - Em_n(x)]$ follows the law of the iterated logarithm. We demonstrate that

$$\limsup_{n\to\infty} \pm [\phi(n)]^{-1} S_n(r(n)) = 1 \quad \text{a.s.},$$

where $\phi(n) = (2n\sigma^2(r)\log\log n)^{1/2}$. An application of this result to the two classes of nonparametric regression function estimators, to be defined below, provides thus a precise description of the order of strong consistency of $m_n(x)$.

The set $\{S_n(r),\, n \geq 1\}$ is a triangular sequence and in this section it is seen that S_n may be strongly approximated by a Gaussian sequence with the same covariance structure. The law of the iterated logarithm will then be shown using parallel results on density estimation by Hall (1981) and Csörgő and Hall (1982). We shall also make use of the Rosenblatt transformation (Rosenblatt, 1952)

$$T(x, y) = (F_{Y|X}, F_X)(x, y),$$

transforming the original data points $\{(X_i, Y_i)\}_{i=1}^n$ into a sequence of mutually independent uniformly distributed over $[0, 1]^2$ random variables $\{(X_i', Y_i')\}_{i=1}^n$. This transformation was also employed by Johnston (1982) as an intermediate tool; also by Mack and Silverman (1982) to obtain (strong) uniform consistency of the Nadaraya-Watson kernel type regression function estimates. It will be convenient to define

$$v_n(u_n) = \int_{|x| \leq u_n} |\, dK_{r(n)}(x) \,| + |\, K_{r(n)}(-u_n -) \,|, \quad n \geq 1$$

with a sequence of constants $\{u_n\}$, $0 < u_n \leq \infty$.

THEOREM 1. *Suppose that the sequence of kernels $K_{r(n)}$ and $\{u_n\}$ satisfy*

$$a_n v_n(u_n) = o(n^{1/2}\sigma(r)(\log\log n)^{1/2}/(\log n)^2), \tag{2.1}$$

where $\{a_n\}$ is a sequence of positive constants tending to infinity. In addition, assume that the following holds.

$$\textstyle\sum_{n=3}^{\infty} E\{K_r^2(X) I(|X| > u_n)\}/(\sigma^2(r)\log\log n) < \infty \tag{2.2a}$$

$$\textstyle\sum_{n=3}^{\infty} E\{K_r^2(X) I(|X| \leq u_n) Y^2 I(|Y| > a_n)\}/(\sigma^2(r)\log\log n) < \infty. \tag{2.2b}$$

Then on a rich enough probability space there exists a Gaussian sequence $\{T_n\}$ with zero means and the same covariance structure as $\{S_n(r)\}$, such that

$$S_n(r) - T_n = o(n^{1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.}$$

The device that is used in the proof is the strong uniform approximation of the empirical process by a Brownian bridge. Hall (1981) employs for density estimation in the one dimensional case the results of Komlós, Major and Tusnády (1975). As in Mack and Silverman (1982), we will make use of an analogous result by Tusnády (1977) for the two dimensional case. Note that although the two dimensional case is considered here, the technique can be extended to higher dimensional design variables $\mathbf{X} = (x^{(1)}, \cdots, x^{(d)})$, $d \geq 2$. The assumption, however, will not be compatible with the case considered here since it is still unknown whether the strong approximation of the multivariate empirical process by a multivariate Brownian bridge has a compatible rate as in the one- or two-dimensional case.

The fundamental connection between $S_n(r)$ and its strong approximation by a Gaussian sequence is established by the following lemma. The proof will be clear from Tusnády (1977) and the fact that $n^{1/2}[F_n(T^{-1}(x, y')) - F(T^{-1}(x', y'))]$, $(x', y') \in [0, 1]^2$ is the empirical process of $\{(X_i, Y_i)\}_{i=1}^n$ (Rosenblatt, 1952).

LEMMA 1. *On a rich enough probability space there is a version of a Brownian bridge* $B(x', y')$, $(x', y') \in [0, 1]^2$ *such that*

$$P\{\sup_{x,y} | e_n(x, y) | > (C_1 \log n + u) \log n\} < C_2 e^{-C_3 u}$$

where C_1, C_2, C_3 *are absolute constants and*

$$e_n(x, y) = n[F_n(x, y) - F(x, y)] - n^{1/2} B(T(x, y)).$$

In the following theorem it is now seen that under regularity conditions on the covariances $\sigma(r, s)$ a law of the iterated logarithm (LIL) holds for $m_n(x)$ as defined in (1.1)

THEOREM 2. *Suppose that* (2.1) *and* (2.2a, b) *hold and that*

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} \sup_{m \in \Gamma_{n,\varepsilon}} | \sigma(r(m), r(n))/\sigma^2(r(n)) - 1 | = 0, \tag{2.3}$$

where $\Gamma_{n,\varepsilon} = \{m: | m - n | \le \varepsilon n\}$. *Then*

$$\limsup_{n \to \infty} \pm [\phi(n)]^{-1} S_n(r) = 1 \quad \text{a.s.}$$

3. Proofs. To establish Theorem 1 we set

$$T_n = n^{1/2} \int \int_{-\infty}^{\infty} K_r(x) y \, dB(T(x, y)),$$

$B(x', y')$ being the Brownian Bridge of Lemma 1, and show that the difference

$$R_n = n^{-1}(S_n(r) - T_n) = n^{-1} \int \int K_r(x) y \, de_n(x, y)$$

satisfies

$$R_n = o(n^{-1/2} \sigma(r) (\log \log n)^{1/2}) \quad \text{a.s.} \tag{3.1}$$

Note first that T_n has the covariance structure ascribed to it in Theorem 1. This follows from the fact that the Jacobian $J(x, y)$ of $T(x, y)$ is $J(x, y) = f(x, y)$, the joint density of (X, Y) (Rosenblatt, 1952) and the following lemma, stated without proof.

LEMMA 2. *Let $G_r(x, y) = K_r(x)y$. Then*

$$(Z_1, Z_2) = \left(\int_0^1 \int_0^1 G_{r_1}(T^{-1}(x', y'))\, dB(x', y'),\right.$$
$$\left.\int_0^1 \int_0^1 G_{r_2}(T^{-1}(x', y'))\, dB(x', y')\right)$$

has a bivariate normal distribution with zero means and covariances

$$\operatorname{cov}(Z_1, Z_2) = \int \int K_{r_1}(x) K_{r_2}(x) y^2 f(x, y)\, dx\, dy$$
$$- \left[\int \int K_{r_1}(x) y f(x, y)\, dx\, dy\right]\left[\int \int K_{r_2}(x) y f(x, y)\, dx\, dy\right]$$
$$= \sigma(r_1, r_2).$$

To demonstrate (3.1), we split up the integration regions and obtain

$$|R_n| \le \textstyle\sum_{j=1}^7 R_{j,n}$$

where

$$R_{1,n} = \left| n^{-1} \int_{|x| \le u_n} \int_{|y| \le a_n} K_r(x) y\, de_n(x, y) \right|$$
$$\le 2v_n(u_n) a_n n^{-1} \sup_{x,y} |e_n(x, y)|,$$
$$R_{2,n} = |n^{-1} \textstyle\sum_{i=1}^n R_{i,n}^{(2)}|,$$
$$R_{i,n}^{(2)} = [K_r(X_i) I(|X_i| > u_n) Y_i I(|Y_i| \le a_n)]$$
$$- E[K_r(X) I(|X| > u_n) Y I(|Y| \le a_n)]$$
$$R_{3,n} = |n^{-1} \textstyle\sum_{i=1}^n R_{i,n}^{(3)}|,$$
$$R_{i,n}^{(3)} = [K_r(X_i) I(|X_i| \le u_n) Y_i I(|Y_i| > a_n)]$$
$$- E[K_r(X) I(|X| \le u_n) Y I(|Y| > a_n)]$$
$$R_{4,n} = |n^{-1} \textstyle\sum_{i=1}^n R_{i,n}^{(4)}|,$$
$$R_{i,n}^{(4)} = [K_r(X_i) I(|X_i| > u_n) Y_i I(|Y_i| > a_n)]$$
$$- E[K_r(X) I(|X| > u_n) Y I(|Y| > a_n)],$$
$$R_{5,n} = n^{-1} \left| \int_{|x| > u_n} \int_{|y| \le a_n} K_r(x) y\, dB(T(x, y)) \right|,$$
$$R_{6,n} = n^{-1} \left| \int_{|x| \le u_n} \int_{|y| > a_n} K_r(x) y\, dB(T(x, y)) \right|,$$
$$R_{7,n} = n^{-1} \left| \int_{|x| > u_n} \int_{|y| > a_n} K_r(x) y\, dB(T(x, y)) \right|.$$

From Lemma 1 we deduce that $n^{-1}\sup_{x,y}|e_n(x, y)| = O(n^{-1}(\log n)^2)$ a.s., and so by condition (2.1) we conclude that

$$R_{1,n} = o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \tag{3.2}$$

Next observe that $\{R_{i,n}^{(2)}\}$ $1 \le i \le n$ are independent and identically distributed random variables. We then have by Markov's inequality that for any $\varepsilon > 0$

$$P(n^{-1}|\textstyle\sum_{i=1}^n R_{i,n}^{(2)}| > \varepsilon\cdot\sigma(r)n^{-1/2}\cdot(\log\log n)^{1/2})$$
$$\le \varepsilon^{-2}\sigma(r)^{-2}(\log\log n)^{-1}\cdot E(R_{1,n}^{(2)})^2.$$

So with the assumption $EY^2 < \infty$ and condition (2.2 a) it follows with the Borel-Cantelli Lemma that

$$R_{2,n} = o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \tag{3.3}$$

The terms $R_{3,n}$, $R_{4,n}$ may be estimated in the same way using Markov's inequality and condition (2.2b) and we therefore have

$$\begin{aligned} R_{3,n} &= o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \\ R_{4,n} &= o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \end{aligned} \tag{3.4}$$

The remaining terms, $R_{5,n}$, $R_{6,n}$ and $R_{7,n}$ are all Gaussian with mean zero and standard deviations

$$\{E(R_{1,n}^{(2)})^2\}^{1/2}, \quad \{E(R_{1,n}^{(3)})^2\}^{1/2}, \quad \{E(R_{1,n}^{(4)})^2\}^{1/2}$$

respectively. Therefore, $R_{5,n}$, for instance, can be computed by

$$P(R_{5,n} > \varepsilon n^{-1/2}\sigma(r)(\log\log n)^{1/2})$$
$$= 2[1 - \Phi\{\varepsilon\sigma(r)(\log\log n)^{1/2}/[E(R_{1,n}^{(2)})^2]^{1/2}\}],$$

where Φ denotes the cdf of the standard normal distribution. A similar equality holds for $R_{6,n}$ and $R_{7,n}$; therefore, we conclude in view of condition (2.2a, b) and the usual approximations to the tails of the normal distribution that

$$\begin{aligned} R_{5,n} &= o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \\ R_{6,n} &= o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \\ R_{7,n} &= o(n^{-1/2}\sigma(r)(\log\log n)^{1/2}) \quad \text{a.s.} \end{aligned} \tag{3.5}$$

Theorem 1 follows now by putting together statements (3.2)–(3.5) respectively.

The proof of Theorem 2 follows in the same way as the proof of Theorem 1 in Hall (1981, page 49). We only have to note that Lemma 1 in Hall (1981, page 49) has to be replaced by (2.3).

4. Kernel estimators. Two types of kernel estimates of the regression function $m(x)$ will be considered here. The first is due to Nadaraya (1964) and Watson (1964):

$$m_n^*(x) = (nh)^{-1}\textstyle\sum_{i=1}^n K((x - X_i)/h)Y_i/[(nh)^{-1}\sum_{i=1}^n K((x - X_i)/h)].$$

We may think of applications where the marginal density f_X of X is known to the statistician. It is then appropriate to replace the density estimator in the denominator of m_n^* by the known density f_X. This leads to the following estimate:

$$\bar{m}_n(x) = (nh)^{-1} \sum_{i=1}^n K((x - X_i)/h) Y_i / f_X(x)$$

considered by Johnston (1979, 1982).

Let us define $S^2(x) = E(Y^2 \mid X = x)$, $V^2(x) = S^2(x) - m^2(x)$, and assume that $f_X(x)$, $m(x)$ are twice differentiable and $S^2(x)$ is continuous. We assume further that the kernel $K(\cdot)$ is continuous, has compact support $(-1, 1)$ and that $\int_{-1}^1 uK(u)\,du = 0$. This implies that $v_n(u_n)$ as defined in (2.1) is constant for large u_n. We will make use of the following assumptions:

$$nh^5/\log\log n \to 0 \quad \text{as} \quad n \to \infty \tag{4.1}$$

$$\sum_{n=3}^\infty (h/\log\log n) E[Y^2 I(|Y| > a_n)] < \infty \tag{4.2}$$

where $\{a_n\}$ is as in (2.1), (2.2 a, b) such that

$$\begin{gathered} a_n = o((nh^{-1}\log\log n)^{1/2}/(\log n)^2). \\ \lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{m\in\Gamma_{n,\epsilon}} | h(m)/h(n) - 1 | = 0. \end{gathered} \tag{4.3}$$

We then have the following theorem for $\bar{m}_n(x)$.

THEOREM 3. *Under the assumptions above*

$$\begin{aligned} \limsup_{n\to\infty} \pm [\bar{m}_n(x) - m(x)](nh/2\log\log n)^{1/2} \\ = [S^2(x) \int K^2(u)\,du/f_X(x)]^{1/2} \quad \text{a.s.} \end{aligned}$$

The Nadaraya-Watson estimate follows also a LIL.

THEOREM 4. *Under the assumptions above and* $\sum_{n=1}^\infty n^{-2}h^{-1} < \infty$

$$\begin{aligned} \limsup_{n\to\infty} \pm [m_n^*(x) - m(x)](nh/2\log\log n)^{1/2} \\ = [V^2(x) \int K^2(u)\,du/f_X(x)]^{1/2} \quad \text{a.s.} \end{aligned}$$

Note that the only difference between Theorem 3 and Theorem 4 is the different scaling factor. As was shown by Johnston (1979), $\bar{m}_n(x)$ has asymptotic variance proportional to $S^2(x)$, whereas $m_n^*(x)$ has asymptotic variance $\sim V^2(x)$. Since in general $S^2(x) \ge V^2(x)$, we expect therefore closer asymptotic confidence intervals for $m_n^*(x)$ than for $\bar{m}_n(x)$.

PROOF OF THEOREM 3. We first show that we could center $\bar{m}_n(x)$ around $E\bar{m}_n(x)$. This follows from

$$E\bar{m}_n(x) = f_X(x)^{-1}h^{-1} \int K((x - u)/h) m(u) f_X(u)\,du = m(x) + O(h^2)$$

using the smoothness of $m(\cdot)$ and $f_X(\cdot)$ and the assumptions on the kernel $K(\cdot)$ (Parzen, 1962; Rosenblatt, 1971).

From assumption (4.1) it thus follows that the bias term $(E\bar{m}_n(x) - m(x))$ vanishes of higher order. So it remains to show that

$$\limsup_{n\to\infty} \pm [\hat{m}_n(x) - E\hat{m}_n(x)]/(nh2 \log\log n)^{1/2} = [S^2(x)\cdot f_X(x) \int K^2(u)\,du]^{1/2} \quad \text{a.s.} \tag{4.4}$$

where $\hat{m}_n(x) = \sum_{i=1}^n K((x - X_i)/h)Y_i = \sum_{i=1}^n K_h(X_i)Y_i$.

From the assumptions on the kernel $K(\cdot)$ we conclude that $\delta_n(u) = h^{-1}K(u/h)$ is a delta function sequence (DFS) in the sense of Watson and Leadbetter (1964). We now make use of this general approach in terms of DFS's and obtain the following:

$$h\sigma^2(h) = h \int \delta_n^2(x - u)S^2(u)f_X(u)\,du - h\left[\int \delta_n(x - u)m(u)f_X(u)\,du\right]^2$$
$$\to S^2(x)\cdot f_X(x) \int K^2(u)\,du \quad \text{as} \quad n \to \infty.$$

This follows from Watson and Leadbetter (1964) by noting that $S^2(\cdot)f_X(\cdot)$ is continuous and $\{h(\int K^2)^{-1}\delta_n^2(u)\}$ is itself a DFS. We may note that the use of this DFS-technique would also provide a slight simplification of Hall's proof (1981) for Rosenblatt-Parzen kernel density estimates.

To establish (4.4) with the use of Theorem 2, we have to show that (2.3) holds. We thus have to demonstrate that if $h, k \to 0$ such that $h/k \to 1$ (in view of assumption (4.3)), then

$$h^{-1}\text{cov}\{K((x - X)/h)Y, K((x - Y)/k)Y\} \to 1. \tag{4.5}$$

But $EK((x - X)/h)Y = h \int \delta_n(x - u)m(u)\cdot f_X(u)\,du = o(h^{1/2})$, and so by the computations for $\sigma^2(h)$ above it remains to demonstrate that

$$h^{-1} \int [K((x - u)/h) - K((x - u)/k)]^2 S^2(u)f_X(u)\,du \to 0.$$

From the boundedness of $S^2(\cdot)$ and $f_X(\cdot)$ it is clear that the integral above is dominated by

$$M \int [K(u) - K(uh/k)]^2\,du.$$

The kernel K is continuous and so $K(uh/k) \to K(u)$ a.e. and it follows that (4.5) holds.

Assumption (2.1) follows from (4.2) since $K(\cdot)$ has compact support and thus $v_n(u_n) = \text{const.}$ for n large enough. In view of the asymptotic formula for $\sigma^2(h)$ above we have by assumption (4.2)

$$a_n = o((n\sigma^2(h)\log\log n)^{1/2}/(\log n)^2)$$

which is assumption (2.1). Finally, assumptions (2.2a, b) follow immediately from (4.2) since K has compact support and as above $\sigma^2(h) \sim h^{-1}$. Theorem 3 thus follows from Theorem 2.

PROOF OF THEOREM 4. To prove Theorem 4 we decompose

$$m_n^*(x) - m(x) = [(nh)^{-1}\hat{m}_n(x) - m(x)f_n(x)]/f_X(x) + f_X^{-1}(x)[m_n^*(x) - m(x)]\cdot[f_X(x) - f_n(x)]$$

where $f_n(x) = (nh)^{-1}\sum_{i=1}^n K((x - X_i)/h)$ is a density estimate of $f_X(x)$. Now from Hall (1981), Theorem 2 it follows that

$$\limsup_{n\to\infty} \pm [f_n(x) - f_X(x)](nh/2 \log\log n)^{1/2} = [f_X(x) \int K^2(u)\,du]^{1/2} \quad \text{a.s.} \tag{4.6}$$

if we use assumption (4.1), ensuring that the bias $(Ef_n(x) - f_X(x)) = O(h^2)$. From Noda (1976) we conclude that with $\sum n^{-2}h^{-1} < \infty$, $m_n^*(x) - m(x) = o(1)$ a.s. This and (4.6) thus yield that the second term on the RHS of the decomposition above is of order $o((nh/2 \log\log n)^{1/2})$ a.s.

The first summand of the decomposition above can be written as

$$\frac{(nh)^{-1}(\hat{m} - E\hat{m})}{f_X} + \frac{(nh)^{-1}E\hat{m} - mf_X}{f_X} - \frac{m(f_n - Ef_n)}{f_X} + \frac{m(f_X - Ef_n)}{f_X}.$$

As in the proof of Theorem 3, it follows by assumption (4.1) that the bias terms $((nh)^{-1}E\hat{m} - mf_X)$ and $(Ef_n - f_X)$ vanish. It remains to show

$$(nh)^{-1}(\hat{m} - E\hat{m}) - m(f_n - Ef_n) \tag{4.7}$$

follows the LIL, i.e.

$$\limsup_{n\to\infty} \pm [(nh)^{-1}(\hat{m} - E\hat{m}) - m(f_n - Ef_n)](nh/2 \log\log n)^{1/2} = [V^2(x)\cdot f_X(x)\cdot \int K^2(u)\,du]^{1/2} \quad \text{a.s.}$$

This can be deduced from Theorem 2, if we rewrite (4.7) as

$$(nh)^{-1} \sum_{i=1}^n [K_h(X_i)Y_i - EK_h(X)Y] - m(x)(nh)^{-1} \sum_{i=1}^n [K_h(X_i) - EK_h(X)] = (nh)^{-1} \sum_{i=1}^n \{K_h(X_i)[Y_i - m(x)] - EK_h(X)[Y - m(x)]\}.$$

Next we show that (4.3) holds. The variance for the sequence above is now:

$$h\cdot\sigma^2(h) = h\cdot\int \delta_n^2(x - u)[S^2(u) - m^2(x)]f_X(u)\,du - h\left[\int \delta_n(x - u)[m(u) - m(x)]f_X(u)\,du\right]^2 \to V^2(x)\cdot f_X(x) \int K^2(u)\,du \quad \text{as } n \to \infty.$$

As above in the proof of Theorem 3, we conclude that (2.3) holds. Theorem 4 thus follows from Theorem 2.

5. Orthogonal polynomial estimators. Estimators of the regression function $m(x)$ based on orthogonal polynomials fit also in the general framework developed in the first section. We define the estimate based on a system of orthonormal polynomials on $[-1, 1]$ as follows:

$$\tilde{m}_n(x) = n^{-1} \textstyle\sum_{i=1}^n K_m(x; X_i) Y_i / n^{-1} \sum_{i=1}^n K_m(x; X_i)$$

where $m = m(n)$ tends with n to infinity and

$$K_m(x; X_i) = \textstyle\sum_{j=0}^m e_j(x) e_j(X_i)$$

and $\{e_j(\cdot)\}$ is the orthonormal system of polynomials. As a technical more tractable estimator we consider also:

$$m_n'(x) = n^{-1} \textstyle\sum_{i=1}^n K_m(x; X_i) Y_i / f_X(x).$$

As in Section 4, let $S^2(x)$ be the second conditional moment of Y and $V^2(x)$ the conditional variance respectively. We further assume that

$$f_X(x) \text{ has compact support in } (-1, 1)$$

$$(1 - x^2)^{-1/4} f_X(x) \text{ is integrable on } (-1, 1).$$

For reasons of simplicity we only consider the case of $e_j(\cdot) = p_j(\cdot) =$ orthonormal Legendre polynomials here and assume that the following holds:

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_{p\in\Gamma_{n,\varepsilon}} |m(p)/m(n) - 1| = 0 \tag{5.1}$$

$$\textstyle\sum_{n=3}^\infty m^{-1} \cdot (\log\log n)^{-1} E(Y^2 \cdot I(|Y| > a_n)) < \infty, \tag{5.2}$$

when $\{a_n\}$ is as in (2.2), (4.2) a sequence of constants tending to infinity such that

$$\begin{gathered} a_n = o(n^{1/2} m (\log\log n)^{1/2} / (\log n)^2). \\ n/(m^5 \log\log n) \to 0 \quad \text{as} \quad n \to \infty. \end{gathered} \tag{5.3}$$

We have then the following Theorem for $m_n'(x)$ and $\tilde{m}_n(x)$.

THEOREM 5. *Under the assumptions above*

$$\limsup_{n\to\infty} \pm [m_n'(x) - m(x)](n/2m \log\log n)^{1/2} = [S^2(x)/(f_X(x)\cdot\pi)]^{1/2} (1 - x^2)^{-1/4} \quad \text{a.s.}$$

and

$$\limsup_{n\to\infty} \pm [\tilde{m}_n(x) - m(x)](n/2m \log\log n)^{1/2} = [V^2(x)/(f_X(x)\cdot\pi)]^{1/2} (1 - x^2)^{-1/4} \quad \text{a.s.}$$

PROOF. We first show the LIL for $m_n'(x)$. The second assertion will then follow as Theorem 4 from Theorem 3. As in Theorem 3, we show first that the bias $(Em_n'(x) - m(x))$ is negligible.

$$\begin{aligned} Em_n'(x) &= [f_X(x)]^{-1} \cdot EK_m(x; X)Y \\ &= [f_X(x)]^{-1} \int K_m(x; u)m(u)f_X(u)d \\ &= m(x) + O(m^{-2}) \end{aligned}$$

by a slight modification of the argument proving Theorem 1 in Walter and Blum (1979). By the same arguments as in Hall's (1981) proof of his Theorem 3 (page 60) we conclude that

$$\sigma_m^2 \sim E[K_m^2(x; X)Y^2] \sim m \cdot S^2(x)/([f_X(x)\pi](1 - x^2)^{1/2}).$$

Assumption (2.1) follows now from (5.2) and

$$\int |dK_m(x; u)| = O(m^2).$$

Assumption (2.2) follows also from (5.2) so we finally derive the desired result from Theorem 2, since (2.3) may be proved as in Theorem 3 using (5.1).

REMARK. There is a wide variety of density estimators based on trigonometric series or Fourier transforms. In the same way as orthogonal polynomial regression estimators are deduced from orthogonal polynomial density estimators, one may construct regression estimators based on trigonometric series. It may be possible to show a law of the iterated logarithm for trigonometric series estimators, but as is indicated in Hall (1981) the computations may be more tedious than for the two classes that are considered here.

Acknowledgement. The author wishes to thank R. Carroll, S. Cambanis and an anonymous referee for helpful suggestions and remarks.

REFERENCES

ČENCOV, N. N. (1962). Evaluation of an unknown distribution density from observations. *Soviet. Math.* **3** 1559–1562.

COLLOMB, G. (1979). Conditions nécessaires et suffisantes de convergence uniform d'un estimateur de la régression, estimation des dérivées de la regression. *C.R. Acad. Sci. Paris* **288** 161–164.

COLLOMB, G. (1981). Estimation non-parametrique de la Regréssion: Revue Bibliographique. *Internat. Statist. Rev.* **49** 75–93.

CSÖRGŐ, S. and HALL, P. (1982). Upper and lower classes for triangular arrays. *Z. Wahrsch. verw. Gebiete* **61** 207–222.

DEVROYE, L. P. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Can. J. Statist.* **6** 179–191.

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

DEVROYE, L. P. and WAGNER, T. J. (1980a). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239.

DEVROYE, L. P. and WAGNER, T. J. (1980b). On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. *Z. Wahrsch. verw. Gebiete* **51** 15–25.

HALL, P. (1981). Laws of the iterated logarithm for nonparametric density estimators. *Z. Wahrsch. verw. Gebiete* **56** 47–61.

JOHNSTON, G. (1979). Smooth nonparametric regression analysis. Inst. of Stat. Mimeo Series No. 1253, University of North Carolina.

JOHNSTON, G. (1982). Probabilities of maximal deviation of nonparametric regression function estimation. *J. Multivariate Anal.* **12** 402–414.

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.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. verw. Gebiete* **61** 405–415.

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

NODA, K. (1976). Estimation of a regression function by the Parzen kernel type density estimators. *Ann. Inst. Math. Statist.* **28** 221–234.

PARZEN, E. (1962). On estimation of a probability density function. *Ann. Math. Statist.* **33** 1065–1076.

ROSENBLATT, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statist.* **23** 470–472.

ROSENBLATT, M. (1969). Conditional probability density and regression estimates. In *Multivariate Analysis* II. 23–51. Ed. Krishnaiah.

RÉVÉSZ, P. (1979). On the nonparametric estimation of the regression function. *Prob. Control. Inform. Theory* **8** 297–302.

RUTKOWSKI, L. (1982a). On system identification by nonparametric function fitting. *IEEE Trans. Int. Control* **27** 225–227.

RUTKOWSKI, L. (1982b). On-line identification of time-varying systems by nonparametric techniques. *IEEE Trans. Int. Control* **27** 228–230.

SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88.

SCHUSTER, E. F. and YAKOWITZ, S. (1979). Contributions to the theory of nonparametric regression, with application to system identification. *Ann. Statist.* **7** 139–149.

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

STUTE, W. (1982). A law of logarithm for kernel density estimators. *Ann. Probab.* **10** 414–422.

TUSNÁDY, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Period. Math. Hung.* **8** 53–55.

WALTER, G. and BLUM, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.* **7** 328–340.

WANDL, H. (1980). On kernel estimation of regression functions. *Wiss. Sitz. zur Stochastik.* **WSS-03** 1–25.

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

WATSON, G. S. and LEADBETTER, M. R. (1964). Hazard analysis II. *Sankhyā* A **26** 101–116.

WEGMAN, E. J. and DAVIES, H. I. (1979). Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7** 316–327.

UNIVERSITÄT HEIDELBERG
SONDERFORSCHUNGSBEREICH 123
IM NEUENHEIMER FELD 293
D-6900 HEIDELBERG 1
WEST GERMANY