A LAW OF THE ITERATED LOGARITHM FOR NONPARAMETRIC REGRESSION FUNCTION ESTIMATORS*

Wolfgang Härdle

Universität Heidelberg
Sonderforschungsbereich 123
Im Neuenheimer Feld 293
D-6900 Heidelberg 1

University of North Carolina
Department of Statistics
321 Phillips Hall 039A
Chapel Hill, NC 27514

Summary: We prove a law of the iterated logarithm for nonparametric regression function estimators using strong approximations to the two dimensional empirical process. We consider the case of Nadaraya-Watson kernel estimators and of estimators based on orthogonal polynomials when the marginal density of the design variable X is unknown or known.

Keywords and Phrases: Nonparametric regression function estimation, law of the iterated logarithm, kernel estimation.

AMS Subject Classifications: 60F10, 60G15, 62G05

*Research partially supported by the Air Force of Scientific Research Contract AFOSR-F49620 82 C 0009.

*Research partially supported by the "Deutsche Forschungsgemeinschaft" SFB123, "Stochastische Mathematische Modelle."

1. Introduction and Background

A series of papers recently appeared on consistency of nonparametric regression function estimators and rates of consistency. (See Collomb, 1981 for a bibliographic review). In the present work we obtain pointwise rates of consistency by demonstrating a law of the iterated logarithm for a large class of regression function estimators. The estimators we shall look at are of the following type:

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

where $\{K_r:\ r \in I\}$ denotes a sequence of delta functions (or kernel sequence) and $\{(X_i, Y_i)\}$ $i=1,2,\ldots,n$ are independent observations of a distribution with unknown positive density $f(x,y)$.

Most nonparametric estimators of $m(x) = E(Y|X=x)$ are of this form, for instance, the Nadaraya-Watson kernel estimator (more generally delta function estimators) or orthogonal polynomial estimators.

A major result in the theory of consistency of kernel type estimators has been obtained by Collomb who gave necessary and sufficient conditions for consistency of the Nadaraya-Watson kernel estimate. For generalizations and related work see the bibliographic review of Collomb (1981) where parallel work on orthogonal polynomials is also presented. Stone (1977) considered the estimator defined in (1.1) and gave general conditions on the weights $K_r(x;\ X_i)$ for $m_n(x)$ to be consistent in L^r, i.e. for

$$E|\hat{m}_n(x) - m(x)|^r \to 0$$

whenever $E|Y|^r < \infty$. Stone, however, points out that it is not clear from his results when an estimator of the Nadaraya-Watson type, to be discussed in section 4, is consistent in L^r. In the field of density estimation Wegman and Davies (1979), Hall (1981), Csörgő and Hall (1982) have given a law of the iterated logarithm for different kinds of density estimators.

We begin by showing a law of the iterated logarithm for the shifted estimate

(1.2) $$m_n(x) - Em_n(x) .$$

That is, we center $m_n(x)$ around its expectation. We could also center it around $m(x)$, the regression curve, but since the bias is purely analytically handled, it suffices to look at (1.2). The handling with these bias terms using different smoothness assumptions of $m(\cdot)$ and $K_r(\cdot)$ 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 estimate with known and unknown marginal density $f_X(x)$ of X and in section 5 we show a similar result for estimators based on orthogonal polynomials.

2. A law of the iterated logarithm for a special triangular array.

Let $\{(X_i,Y_i)\}$ be a sequence of independent and identically distributed rv's with pdf $f(x,y)$ and cdf $F(x,y)$ 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

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

which is actually a multiple of (1.2) where we omitted 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 establish conditions similar to Hall (1981) and Csörgő and Hall (1982) under which $S_n(r)$, $r=r(n)\in I$ follows the law of the iterated logarithm. We demonstrate that

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

where $\phi(n) = (2n\sigma^2(r)\log\log n)^{1/2}$. The set $\{S_n(r), n \geq 1\}$ is, in fact, a triangular sequence, and in this section it is shown that under certain assumptions S_n may be approximated by a Gaussian sequence with the same covariance structure. A law of the iterated logarithm can then easily be deduced using techniques similar to Hall (1981).

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) and Mack and Silverman (1982) to obtain strong uniform consistency of Nadaraya-Watson kernel type regression function estimates. Define

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

where $\{u_n\}$ is a sequence of constants $0<u_n\leq\infty$.

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

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

where $\{a_n\}$ is a sequence of positive constants tending to infinity.

$$\sum_{n=3}^{\infty} \sigma(r)^{-2}(\log\log n)^{-1}[E\{K_r^2(X)\cdot I(|X|>u_n)\}] < \infty$$

(2.2) $$\sum_{n=3}^{\infty} \sigma(r)^{-2}(\log\log n)^{-1}[E\{K_r^2(X)\cdot I(|X|\leq u_n)\cdot Y^2\cdot I(|Y|>a_n)\}] < \infty .$$

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

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

The main idea of the proof is as in Hall (1981) (for density estimators) and in Härdle (1983) (for regression estimators) the strong approximation of $F_n(z)-F(z)$ (density case) and of $F_n(x,y)-F(x,y)$ (regression case) respectively. Hall employs for the case of density estimation the results of Komlós, Major, Tusnády, (1975). We will make use of a similar result (for the two dimensional case) by Tusnády (1977). The fundamental connection between the regression estimator $\hat{m}_n(\cdot)$ and its strong approximation by a Gaussian process is established by the following lemma.

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

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

where C_1, C_2, C_3 are absolute constants and

$$e_n(x,y) = n[(F_n(x,y)-F(x,y))-B(T(x,y))] .$$

Proof. This is 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).

The following theorem establishes now under regularity conditions on the covariance matrix $\sigma(r,s)$ that a law of the iterated logarithm (LIL) holds for $\hat{m}_n(x)$ the regression function estimator as defined in (1.1).

Theorem 2. Suppose that (2.1) and (2.2) hold and that

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

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

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

Condition (2.3) is the same as in Hall (1981) but with $\sigma(r_1,r_2) = \iint y^2 K_{r_1}(x)K_{r_2}(x)f(x,y)dxdy$ instead of his $\sigma_{r_1,r_2} = \int K_{r_1}(x)K_{r_2}(x)dx$ in the case of estimating the uniform density.

3. Proofs.

To establish Theorem 1 we set

$$T_n = n\iint_{-\infty}^{\infty} K_r(x)ydB(T(x,y)) \quad ,$$

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

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

satisfies

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

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) (see Rosenblatt, 1952) and the following lemma, stated without proof.

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

$$(Z_1,Z_2) = \left(\iint_{00}^{11} G_{r_1}(T^{-1}(x',y'))dB(x',y') \; , \; \iint_{00}^{11} G_{r_2}(T^{-1}(x',y'))dB(x',y')\right)$$

has a bivariate normal distribution with zero means and covariances

$$\begin{aligned} \mathrm{cov}(Z_1,Z_2) &= \iint K_{r_1}(x)K_{r_2}(x)y^2 f(x,y)dxdy \\ &\quad - [\iint K_{r_1}(x)yf(x,y)dxdy][\iint K_{r_2}(x)yf(x,y)dxdy] \\ &= \sigma(r_1,r_2) \; . \end{aligned}$$

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

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

where

$$R_{1,n} = \left| n^{-1} \int_{|x|\le u_n} \int_{|y|\le a_n} K_r(x) y\, de_n(x,y) \right|$$

$$\le v_n(u_n)\cdot 2\cdot a_n \cdot n^{-1}\cdot \sup_{x,y} |e_n(x,y)| \ ,$$

$$R_{2,n} = \left| n^{-1} \sum_{i=1}^{n} R_{i,n}^{(2)} \right| \ ,$$

$$R_{i,n}^{(2)} = [K_r(X_i)\cdot I(|X_i|>u_n)\cdot Y_i \cdot I(|Y_i|\le a_n)]$$

$$- E[K_r(X) I(|X|>u_n) Y\cdot I(|Y|\le a_n)]$$

$$R_{3,n} = \left| n^{-1} \sum_{i=1}^{n} R_{i,n}^{(3)} \right| \ ,$$

$$R_{i,n}^{(3)} = [K_r(X_i)\cdot I(|X_i|\le u_n)\cdot Y_i \cdot I(|Y_i|>a_n)]$$

$$- E[K_r(X)\cdot I(|X|\le u_n)\cdot Y\cdot I(|Y|>a_n)]$$

$$R_{4,n} = \left| n^{-1} \sum_{i=1}^{n} R_{i,n}^{(4)} \right| \ ,$$

$$R_{i,n}^{(4)} = [K_r(X_i)\cdot I(|X_i|>u_n)\cdot Y_i \cdot I(|Y_i|>a_n)]$$

$$- E[K_r(X)\cdot I(|X|>u_n)\cdot Y\cdot I(|Y|>a_n)] \ ,$$

$$R_{5,n} = n^{-1} \left| \int_{|x|>u_n} \int_{|y|\le a_n} K_r(x) y\, dB(T(x,y)) \right| \ ,$$

$$R_{6,n} = n^{-1} \left| \int_{|x|\le u_n} \int_{|y|>a_n} K_r(x) y\, dB(T(x,y)) \right| \ ,$$

$$R_{7,n} = n^{-1} \left| \int_{|x|>u_n} \int_{|y|>a_n} K_r(x) y\, dB(T(x,y)) \right| \ .$$

From Lemma 1 we deduce that $n^{-1}\sup_{x,y}|e_n(x,y)| = 0(n^{-1}(\log n)^2)$ 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 $\{R^{(2)}_{i,n}\}$ $1 \le i \le n$ are independent and identically distributed random variables. We then have by Markov's inequality that for any $\varepsilon>0$

$$P(n^{-1}|\sum_{i=1}^{n} R^{(2)}_{i,n}|>\varepsilon\cdot\sigma(r)n^{-1/2}\cdot(\log\log n)^{1/2})$$

$$\le \varepsilon^{-2}\sigma(r)^{-2}(\log\log n)^{-1}\cdot E(R^{(2)}_{1,n})^{2} \ .$$

So with the assumption $EY^2<\infty$ and condition (2.2) 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 (2.2) 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

$$\{E(R^{(2)}_{1,n})^{2}\}^{1/2}$$

$$\{E(R^{(3)}_{1,n})^{2}\}^{1/2}$$

$$\{E(R^{(4)}_{1,n})^{2}\}^{1/2}$$

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

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

$$= 2[1-\Phi\{\varepsilon\cdot\sigma(r)\cdot(\log\log n)^{1/2}/[E(R^{(2)}_{1,n})^{2}]^{1/2}\}] \ ,$$

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 (2.2) and the usual approximations to the tails of the normal distribution that

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

Finally the desired result of Theorem 1 follows by putting together statements (3.2)-(3.5) respectively.

The proof of Theorem 2 follows in much the same way as Theorem 1 in Hall (1981, p. 49). We only have to note that lemma 1 in Hall (1981, p. 49) has to be replaced by (2.3). Setting $Y \equiv 1$ in all our derivations shows that Hall's result follows from ours.

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) and is motivated by the formula

$$m(x) = \{\int y f(x,y)\,dy\}/f_X(x) .$$

We define the Nadaraya-Watson estimate as follows:

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

Consistency and asymptotic normality of $m_n^*(x)$ were considered by Schuster (1972), Johnston (1979), Mack and Silverman (1982) among others. If the marginal density $f_X(x)$ is known, it is appropriate to replace the density estimate in the denominator of $m_n^*(x)$ by the true density $f_X(x)$. 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 $S^2(x) = E(Y^2|X=x)$, $V^2(x) = S^2(x) - m^2(x)$, 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) say and that $\int_{-1}^{1} uK(u)du = 0$. This implies that $v_n(u_n)$ as used in (2.1) is constant for large enough u_n. We will make use of the following assumptions:

(4.1) $\quad nh^5/\text{loglogn} \to 0 \quad \text{as } n \to \infty$

(4.2) $\quad \sum_{n=3}^{\infty} h(\text{loglogn})^{-1} \cdot E[Y^2 I(|Y|>a_n)] < \infty$

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

$$a_n = o((nh^{-1}\text{loglogn})^{1/2}/(\text{logn})^2) \quad .$$

(4.3) $\quad \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_{m\in\Gamma_{n,\varepsilon}} |h(m)/h(n) - 1| = 0 \quad .$

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

Theorem 3. Under the assumptions above

$$\limsup_{n\to\infty} \pm [\bar{m}_n(x) - m(x)](nh/2\text{loglogn})^{1/2}$$

$$= [S^2(x) \int K^2(u)du / f_X(x)]^{1/2} \quad \text{a.s.}$$

The Nadaraya-Watson estimate follows also a LIL as the following theorem shows.

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

$$\limsup_{n\to\infty} \pm [m_n^*(x) - m(x)](nh/2\text{loglogn})^{1/2}$$

$$= [V^2(x) \int K^2(u)du / f_X(x)]^{1/2} \quad \text{a.s.}$$

Note that the only difference between Theorem 3 and Theorem 4 is the different scaling factor. Since in general $S^2(x) \geq V^2(x)$ we may expect closer asymptotic confidence bands for $m_n^*(x)$. This observation has already been made by Schuster (1972) and Johnston (1982). This papers together with Härdle (1983) thus solves the question raised by Johnston (1982) whether one should be able to compute asymptotic confidence intervals for $m_n^*(x)$. Johnston derived (uniform) confidence intervals for $\bar{m}_n(x)$ only.

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

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

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) (Watson and Leadbetter, 1964). We make now use of this general approach in terms of DFS's and obtain the following:

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

This follows from Watson and Leadbetter (1964) by noting that $S^2(\cdot) f_X(\cdot)$ is continuous and $\{h(\int K^2)^{-1}\cdot\delta_n^2(u)\}$ is itself a DFS. The use of this DFS-technique would also considerably simplify 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 must thus demonstrate that if $h,k\to 0$ such that $h/k\to 1$ (in view of assumption (4.3)), then

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

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 $v_n(u_n)$ = 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, assumption (2.2) follows 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

$$\text{(4.6)} \qquad \limsup_{n\to\infty} \pm [f_n(x)-f_X(x)](nh/2\log\log n)^{1/2}$$

$$= [f_X(x)\int K^2(u)du]^{1/2} \quad \text{a.s.}$$

if we use assumption (4.1) which ensures that the bias $(Ef_n(x)-f_X(x)) = O(h^2)$ vanishes. Note that Hall's assumption (11) is not necessary here since we assume that $K(\cdot)$ has compact support. From Noda (1976) we conclude that $\sum n^{-2}h^{-1} < \infty$ makes $m_n^*(x)-m(x)=o(1)$ 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

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

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

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

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:

$$h \cdot \sigma^2(h) = h \cdot \int \delta_u^2(x-u)[S^2(u)-m^2(x)]f_X(u)du$$

$$- h[\int \delta_n(x-u)[m(u)-m(x)]f_X(u)du]^2$$

$$\to V^2(x) \cdot f_X(x) \int K^2(u)du \qquad \text{as } n\to\infty .$$

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:

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

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

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

and $\{e_j(\cdot)\}$ is the orthonormal system of polynomials.

In the case of a known marginal density $f_X(x)$ we consider

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

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

$f_X(x)$ has compact support in $(-1,1)$

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

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

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

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

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

$$a_n = o(n^{1/2} m(\log\log n)^{1/2}/(\log n)^2) .$$

(5.3) $$n/(m^5 \log\log n) \to 0 \quad \text{as} \quad n\to\infty .$$

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

<u>Theorem 5</u>. 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 that 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.

$$Em_n'(x) = [f_X(x)]^{-1}\cdot EK_m(x;\, X)Y$$

$$= [f_X(x)]^{-1}\int K_m(x;\, u)m(u)f_X(u)d$$

$$= m(x) + O(m^{-2})$$

y 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 (p. 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).

Acknowledgement

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

References

1. Collomb, G. (1981). Estimation non-parametrique de la Regrèssion: Revue Bibliographique. International Statistical Review 49, 75-93.

2. Csörgő, S. and Hall, P. (1982). Upper and lower classes for triangular arrays, Z. f. Wahrsch. verw. Geb. 61, 207-222.

3. Hall, P. (1981). Laws of the iterated logarithm for nonparametric density estimators. Z. f. Wahrsch. verw. Geb. 56, 47-61.

4. Härdle, W. (1983). Asymptotic maximal deviation of M-smoothers. Submitted for publication.

5. Johnston, G. (1979). Smooth nonparametric regression analysis. Inst. of Stat. Mimeo Series #1253, University of North Carolina.

6. Johnston, G. (1982). Probabilities of maximal deviation of nonparametric regression function estimation. J. Mult. Anal. 12, 402-414.

7. Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df I. Z. f. Wahrsch. verw. Geb. 32, 111-131.

8. Mack, Y.P. and Silverman, B.W. (1982). Weak and strong uniform consistency of kernel regression estimates. Z. F. Wahrsch. verw. Geb. 61, 405-415.

8a. Nadaraya, E.A. (1964). On estimating regression. Theor. Prob. Appl. 9, 141-142.

9. Noda, K. (1976). Estimation of a regression function by the Parzen kernel type density estimators. Inst. Stat. Math. , 221-234.

10. Parzen, E. (1962). On estimation of a probability density function. Ann. Math. Stat. 33, 1065-1076.

11. Rosenblatt, M. (1952). Remarks on a multivariate transformation, Ann. Math. Stat. 23, 470-472.

12. Rosenblatt, M. (1971). Curve estimates. Ann. Math. Stat. 42, 1815-1842.

13. Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. Ann. Math. Stat. 43, 84-88.

14. Stone, C. (1977). Consistent nonparametric regression. Ann. Stat. 5, 595-645.

15. Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. Period. Math. Hung. 8, 53-55.

16. Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences. Ann. Stat. 7, 328-340.

16a. Watson, G.S. (1964). Smooth regression analysis, Sankhyā, Series A, Vol. 26, 359-372.

17. Watson, G.S. and Leadbetter, M.R. (1964). Hazard analysis II. Sankhyā, Series A, 26, 101-116.

18. Wegman, E.J. and Davies, H.I. (1979). Remarks on some recursive estimators of a probability density. Ann. Stat. 7, 316-327.