Z. Wahrscheinlichkeitstheorie verw. Gebiete
56, 47–61 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Laws of the Iterated Logarithm for Nonparametric Density Estimators

Peter Hall

Australian National University, Dept. of Statistics, Box 4, Canberra ACT 2600, Australia

Summary. We establish a law of the iterated logarithm for a triangular array of independent random variables, and apply it to obtain laws for a large class of nonparametric density estimators. We consider the case of Rosenblatt-Parzen kernel estimators, trigonometric series estimators and orthogonal polynomial estimators in detail, and point out that our technique has wider application.

1. Introduction and Summary

Much attention has recently been directed towards obtaining conditions for the strong consistency of density estimators, and rates of strong consistency. In the present paper we obtain sharp pointwise rates of strong consistency, by establishing laws of the iterated logarithm for a large class of estimators. The estimators we shall consider are of the type

$$\hat{f}_n(x)=n^{-1}\sum_{i=1}^{n} K_{r(n)}(x;X_i),$$

where $\{K_r, r\in I\}$ is a sequence of "kernel" functions (I is an arbitrary index set), and $X_1, X_2, \ldots$ are independent observations of a distribution with unknown density f. Most nonparametric estimators have this form – for example, Rosenblatt-Parzen kernel estimators, trigonometric series estimators, orthogonal polynomial estimators, Fourier transform estimators and histogram estimators.

A fundamental contribution to the theory of strong consistency has been made by Deheuvels [7], who gave necessary and sufficient conditions for the strong convergence of Rosenblatt-Parzen kernel estimators. For generalizations and related work see the more recent papers of Devroye [8], Devroye and Wagner [9], Silverman [14] and Singh [15]. Earlier work on kernel estimators is referenced in these articles. Parallel results on the convergence of orthogonal

0044-3719/81/0056/0047/$03.00

series estimators have been obtained by Bluez and Bosq [1, 2] and Bosq [3, 4, 5]. See also Winter [19]. Wegman and Davies [18] have given a law of the iterated logarithm for a sequentially calculated density estimator, but their results do not overlap with ours.

We begin by establishing a law of the iterated logarithm for a general class of triangular arrays of independent variables, complementing work of Tomkins [17]. This result is presented in Section 2, and is central to the paper. In the subsequent sections we apply it to establish laws for Rosenblatt-Parzen kernel estimators, trigonometric series estimators and orthogonal polynomial estimators. The special properties of the "kernel" functions used to construct these estimators necessitate some differences of detail in the proofs, but there are many similarities and we keep our discussion brief. We stress that these three types of estimators are only intended as examples, and our method can be applied more widely.

The rate of convergence of a density estimator $\hat{f}_n(x)$ to the density $f(x)$ is restricted by two different factors – the rate of convergence of the error from the mean, $\hat{f}_n(x)-E[\hat{f}_n(x)]$, and the rate of convergence of the bias, $E[\hat{f}_n(x)]-f(x)$. Generally these factors work against one another, in the sense that a construction which reduces one will tend to increase the other. Either term may dominate the difference $\hat{f}_n(x)-f(x)$. Since the bias is purely deterministic and may be easily estimated by analytical methods, it suffices to obtain a law of the iterated logarithm for the error from the mean. We shall adopt this procedure.

In the proofs the symbol C will denote a positive generic constant not depending on the parameters in question. It will differ from appearance to appearance.

2. A Law of the Iterated Logarithm for Triangular Arrays

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with their distribution confined to the interval (a, b), where $-\infty \leqq a < b \leqq \infty$. Let $\{K_r,\ r \in I\}$ be a sequence of univariate functions each of bounded variation on (a, b), and define

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

$$\sigma_{rs} = \operatorname{cov}[K_r(X_1), K_s(X_1)] \quad \text{and} \quad \sigma_r^2 = \sigma_{rr}.$$

Select a sequence $\{r(n),\ n=1,2,\ldots\} \subseteq I$. We establish conditions under which $S_n(r(n))$ obeys the law of the iterated logarithm. That is, with

$$\phi(n) = (2n\sigma_{r(n)}^2 \log\log n)^{\frac{1}{2}}$$

we have

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

almost surely (a.s.). We shall write S_n for $S_n(r(n))$ and r for $r(n)$ whenever no ambiguity arises, and denote an integral over (a, b) by $\int$.

Theorem 1. *Suppose*

$$(\log n)^4 \left[\int |dK_{r(n)}(x)|\right]^2 / n\sigma^2_{r(n)} \log\log n \to 0 \tag{2}$$

and

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_m |\sigma_{r(m)\,r(n)}/\sigma^2_{r(n)} - 1| = 0, \tag{3}$$

where the inner supremum is taken over values of m with $|m-n| \leqq \varepsilon n$. Then (1) *holds.*

Condition (3) may be reformulated as: for all sequences $m, n \to \infty$ with $m/n \to 1$,

$$\sigma_{r(m)\,r(n)}/\sigma^2_{r(n)} \to 1.$$

With $K_r(x) = x$ for each r we obtain as a corollary the Hartman-Wintner law for a sequence of bounded variables. Next suppose

$$K_r(x) = \begin{cases} 1 & \text{if } x \leqq r, \\ 0 & \text{if } x > r, \end{cases}$$

where $0 \leqq r \leqq 1$, and $X_1, X_2, \ldots$ have the uniform distribution on $(0,1)$. Let $r(n) \to 0$,

$$(\log n)^4 / n r(n) \log\log n \to 0, \tag{4}$$

and suppose $r(m)/r(n) \to 1$ for all sequences $m, n \to \infty$ such that $m/n \to 1$. Then as a corollary to Theorem 1,

$$\limsup_{n\to\infty} \pm (2nr \log\log n)^{-\frac{1}{2}} n[F_n(r) - r] = 1 \quad \text{a.s.},$$

where F_n is the empiric distribution function of $X_1, X_2, \ldots, X_n$, and $r = r(n)$. Eicker [10] and Kiefer [11] obtained this result by other methods, and Kiefer showed that (4) may be weakened to

$$\log\log n / n r(n) \to 0.$$

Tomkins [17] considered the law of the iterated logarithm for triangular arrays in a more general setting, but only obtained a lower bound. For a general sequence his conditions would be difficult to check.

Proof of Theorem 1. We begin with

Lemma 1. *Let L_1 and L_2 be functions of bounded variation on* $(0,1)$, *and $W^0(t)$, $0 \leqq t \leqq 1$, be a Brownian bridge. Then*

$$(Z_1, Z_2) = \left(\int_0^1 W^0(t)\, dL_1(t),\ \int_0^1 W^0(t)\, dL_2(t)\right)$$

has a bivariate normal distribution with zero means and covariances

$$\operatorname{cov}(Z_i, Z_j) = \int_0^1 L_i(t)\, L_j(t)\, dt - \left[\int_0^1 L_i(t)\, dt\right] \left[\int_0^1 L_j(t)\, dt\right].$$

The proof is straightforward.

Returning to the proof of Theorem 1 we see that it suffices to consider the case where each X_i is uniform on $(0,1)$. In this case $a=0$ and $b=1$. Let F_n denote the empiric distribution function of $X_1, X_2, \ldots, X_n$. In view of Theorem 4 of Komlós, Major and Tusnády [12], on a rich enough probability space we may write

$$n[F_n(x)-x]=\sum_{i=1}^{n} W_i^0(x)+e_n(x),$$

$0 \leqq x \leqq 1$, where W_i^0, $i \geqq 1$, are independent Brownian bridges and there exist positive absolute constants C_1, C_2 and λ such that

$$P(\sup_{0\leqq x\leqq 1} |e_n(x)|>(C_1 \log n+x)\log n)<C_2 e^{-\lambda x}$$

for all x and n. (Komlós, Major and Tusnády speculate that the extra factor of $\log n$ is unnecessary; this would allow us to reduce $(\log n)^4$ to $(\log n)^2$ in condition (2).) Therefore

$$S_n=n\int K_r(x)\,d[F_n(x)-x]=-\sum_{i=1}^{n}\int W_i^0(x)\,dK_r(x)-\int e_n(x)\,dK_r(x).$$

For large values of n, making use of condition (2), we have

$$\begin{aligned}
&P([\phi(n)]^{-1}|\int e_n(x)\,dK_r(x)|>\varepsilon)\\
&\quad\leqq P(\sup_{0\leqq x\leqq 1}|e_n(x)|>\varepsilon(\log n)^2\,[2n\sigma_r^2 \log\log n/(\log n)^4\,\{\int|dK_r(x)|\}^2]^{\frac{1}{2}})\\
&\quad\leqq C_2 \exp(-\delta\,[n\sigma_r^2 \log\log n/(\log n)^2\,\{\int|dK_r(x)|\}^2]^{\frac{1}{2}}),
\end{aligned}$$

where $\delta>0$ does not depend on n. It follows from the Borel-Cantelli lemma that

$$[\phi(n)]^{-1}\int e_n(x)\,dK_r(x)\to 0 \quad \text{a.s.},$$

and so it suffices to prove that the variables

$$T_n(r)=\sum_{i=1}^{n}\int W_i^0(x)\,dK_r(x)$$

obey the law of the iterated logarithm. We shall write T_n for $T_n(r(n))$.

For fixed n we have

$$T_n=n^{\frac{1}{2}}\int W^0(t)\,dK_r(t)$$

for a Brownian bridge W^0. In view of the lemma and the usual approximation to the tail of the normal distribution,

$$P(T_n>(1+\varepsilon)\,\phi(n))\leqq(2\pi)^{-\frac{1}{2}}\exp(-(1+\varepsilon)^2\log\log n). \tag{5}$$

Let $\rho>1$ and set $m_k=[\rho^k]$ (the integer part of ρ^k). We seek a lower bound for

$$\begin{aligned}
&P(T_{m_k}>(1+\varepsilon)\,\phi(m_k)\,|\,T_n>(1+3\varepsilon)\,\phi(n) \text{ for some } m_{k-1}<n\leqq m_k)\\
&\quad\geqq \inf P(T_{m_k}>(1+\varepsilon)\,\phi(m_k)\,|\,T_n=z),
\end{aligned} \tag{6}$$

where the infemum is taken over integers n with $m_{k-1}<n\leqq m_k$ and real numbers $z\geqq(1+3\varepsilon)\phi(n)$. For any $m\geqq n$ the variables T_m and T_n have a joint normal distribution with zero means, variances $m\sigma^2_{r(m)}$ and $n\sigma^2_{r(n)}$ respectively, and covariance $n\sigma_{r(m)r(n)}$ (use Lemma 1). Therefore conditional on $T_n=z, T_m$ is normal

$$N(\sigma_{st}z/\sigma_t^2, m\sigma_s^2-n\sigma_{st}^2/\sigma_t^2)$$

where $s=r(m)$ and $t=r(n)$. Let Φ denote the standard normal distribution function. Then if $\sigma_{st}>0$,

$$\begin{aligned}P(T_m&>(1+\varepsilon)\phi(m)|T_n=z)\\&=1-\Phi[\{(1+\varepsilon)\phi(m)-\sigma_{st}z/\sigma_t^2\}/\{m\sigma_s^2-n\sigma_{st}^2/\sigma_t^2\}^{\frac{1}{2}}]\\&\geqq 1-\Phi[\{(1+\varepsilon)\phi(m)-(1+3\varepsilon)\phi(n)\sigma_{st}/\sigma_t^2\}/\{m\sigma_s^2-n\sigma_{st}^2/\sigma_t^2\}^{\frac{1}{2}}]\end{aligned}$$

if $z\geqq(1+3\varepsilon)\phi(n)$. By choosing ρ sufficiently close to 1 we may ensure that

$$(1+3\varepsilon)(\mathrm{n}\log\log n)^{\frac{1}{2}}\geqq(1+2\varepsilon)(m_k\log\log m_k)^{\frac{1}{2}}$$

for all $m_{k-1}<n\leqq m_k$ and all sufficiently large k. And applying condition (3) we see that if ρ is close to 1,

$$(1+2\varepsilon)\sigma_{r(m_k)r(n)}/\sigma_{r(n)}\geqq(1+\varepsilon)\sigma_{r(m_k)}$$

for all $m_{k-1}<n\leqq m_k$ and all large k. Therefore for large k,

$$P(T_{m_k}>(1+\varepsilon)\phi(m_k)|T_n=z)\geqq 1-\Phi(0)=\tfrac{1}{2}$$

uniformly in $m_{k-1}<n\leqq m_k$ and $z\geqq(1+3\varepsilon)\phi(n)$. From this result, (5), (6) and the inequality

$$P(A)\leqq P(B)/P(B|A)$$

we see that for an integer k_0,

$$\begin{aligned}\sum_{k\geqq k_0}&P(T_n>(1+3\varepsilon)\phi(n)\text{ for some }m_{k-1}<n\leqq m_k)\\&\leqq 2\sum_{k\geqq k_0}P(T_{m_k}>(1+\varepsilon)\phi(m_k))<\infty.\end{aligned}$$

The Borel-Cantelli lemma now implies that for all $\varepsilon>0$,

$$P(T_n>(1+\varepsilon)\phi(n)\,\text{i.o.})=0. \tag{7}$$

A lower bound may be obtained by applying Tomkins' [17] Theorem 1, but it is very easy to give a direct proof. To this end, define

$$\psi(k)=[2(m_k-m_{k-1})\sigma^2_{r(m_k)}\log\log m_k]^{\frac{1}{2}}$$

and

$$\Delta_k=T_{m_k}(r(m_k))-T_{m_{k-1}}(r(m_k)).$$

The variables Δ_k are independent and normally distributed with zero means and variances $(m_k-m_{k-1})\sigma^2_{r(m_k)}$, and so by the Borel-Cantelli lemma and an estimate of the tail of the normal distribution,

$$P(\Delta_k > (1-\varepsilon)\psi(k)\text{ i.o.}) = 1 \tag{8}$$

for all $\varepsilon > 0$. Choose $\rho > 1$ so large that

$$[1-2\varepsilon-(1-\varepsilon)(1-\rho^{-1})^{\frac{1}{2}}]\rho^{\frac{1}{2}} < -3.$$

For large values of k,

$$-\xi(k) = (1-2\varepsilon)\phi(m_k) - (1-\varepsilon)\psi(k) \leqq -2(2m_{k-1}\sigma^2_{r(m_k)}\log\log m_{k-1})^{\frac{1}{2}}.$$

The variable $T_{m_{k-1}}(r(m_k))$ is normally distributed with zero mean and variance $m_{k-1}\sigma^2_{r(m_k)}$, and so by the Borel-Cantelli lemma,

$$P(T_{m_{k-1}}(r(m_k)) \leqq -\xi(k)\text{ i.o.}) = 0.$$

Combining this with (8) we see that for any $\varepsilon > 0$,

$$P(\Delta_k + T_{m_{k-1}}(r(m_k)) > (1-\varepsilon)\psi(k) - \xi(k)\text{ i.o.}) = 1;$$

that is

$$P(T_{m_k}(r(m_k)) > (1-2\varepsilon)\phi(m_k)\text{ i.o.}) = 1.$$

One part of the law of the iterated logarithm is a consequence of this and (7); the other follows by symmetry.

3. Rosenblatt-Parzen Kernel Estimators

Let K be a function of bounded variation on $(-\infty, \infty)$ satisfying

$$zK(z) \to 0 \quad \text{as} \quad |z| \to \infty \quad \text{and} \quad \int_{-\infty}^{\infty} K^2(z)\,dz < \infty. \tag{9}$$

Let $X_1, X_2, \ldots$ be independent random variables whose common distribution function F has a derivative $F'(x) = f(x) \neq 0$ at x. A kernel estimator of $f(x)$ is defined by

$$\hat{f}_n(x) = (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)/h),$$

where $h = h(n)$ is a sequence of positive constants converging to zero.

We shall assume in addition that F satisfies a Lipshitz condition of order one in a neighbourhood of x (that is, for some ε, $M > 0$, $|F(y) - F(z)| \leqq M|y-z|$ whenever $|x-y|$ and $|x-z| < \varepsilon$); that

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\sup_{m} |h(m)/h(n) - 1| = 0, \tag{10}$$

where the inner supremum is taken over values of m with $|m-n| \leqq n\varepsilon$; and that

$$(\log n)^4/nh\log\log n \to 0. \tag{11}$$

Conditions (10) and (11) would be satisfied in practice, since it is usual to take $h(n) \sim a n^{-b}$ for positve numbers a and b with $b<1$. If f exists and is uniformly continuous in a neighbourhood of x, then condition (11) implies the uniform convergence of the estimate $\hat{f}_n$ within a neighbourhood.

From Theorem 1 we obtain

Theorem 2. *Under the conditions above,*

$$\limsup_{n\to\infty} \pm [\hat{f}_n(x) - E\hat{f}_n(x)]\,(nh/2\log\log n)^{\frac{1}{2}} = \left[f(x) \int_{-\infty}^{\infty} K^2(z)\,dz\right]^{\frac{1}{2}} \text{ a.s.}$$

Analogues of theorem 2 can be obtained for $\hat{f}_n^{(p)}$, with stronger hypotheses on f and K, in the estimation of the p'th derivative $f^{(p)}$.

Proof. We set $b = -a = \infty$ and $K_h(y) = K((x-y)/h)$, and verify conditions (2) and (3). Now,

$$\sigma_h^2 = 2\int_{-\infty}^{\infty} [F(x-zh) - F(x)]\,K(z)\,dK(z) - \left\{\int_{-\infty}^{\infty} [F(x-zh) - F(x)]\,dK(z)\right\}^2,$$

and for any $\varepsilon > 0$,

$$\int_{|z|>\varepsilon/h} |[F(x-zh) - F(x)]\,K(z)\,dK(z)| \leqq \varepsilon^{-1} h \int_{|z|>\varepsilon/h} |z K(z)\,dK(z)| = o(h),$$

using (9), while

$$\begin{aligned} &\left|\int_{|z|>\varepsilon/h} [F(x-zh) - F(x)]\,dK(z)\right| \\ &\quad \leqq |K(\varepsilon/h)| + |K(-\varepsilon/h)| + \int_{|x-y|>\varepsilon} |K((x-y)/h)|\,dF(y) \\ &\quad \leqq |K(\varepsilon/h)| + |K(-\varepsilon/h)| \\ &\qquad + \varepsilon^{-1} h \int_{|x-y|>\varepsilon} |h^{-1}(x-y)\,K(h^{-1}(x-y))|\,dF(y) \\ &\quad = o(h). \end{aligned}$$

Since

$$\sup_{0<|u|<\varepsilon} |u^{-1}[F(x+u) - F(x) - u f(x)]| \to 0 \tag{12}$$

as $\varepsilon \to 0$ then for any $\delta > 0$ we may choose ε so small that

$$\left|\int_{|z|\leqq\varepsilon/h} [F(x-zh) - F(x) + zh f(x)]\,K(z)\,dK(z)\right| \leqq \delta h$$

and

$$\left|\int_{|z|\leqq\varepsilon/h} [F(x-zh) - F(x) + zh f(x)]\,dK(z)\right| \leqq \delta h$$

for all h. An integration by parts and the Cauchy-Schwartz inequality give the formula

$$\begin{aligned} \left|\int_{|z|\leqq\varepsilon/h} z\,dK(z)\right| &\leqq (\varepsilon/h)[|K(\varepsilon/h+)| + |K(-\varepsilon/h-)|] \\ &\quad + \left[\int_{|z|\leqq\varepsilon/h} K^2(z)\,dz\right]^{\frac{1}{2}} \left[\int_{|z|\leqq\varepsilon/h} dz\right]^{\frac{1}{2}} = O(h^{-\frac{1}{2}}), \end{aligned}$$

and combining these estimates we deduce that

$$\sigma_h^2 \sim E[K^2((x-X_1)/h)] \sim hf(x) \int_{-\infty}^{\infty} K^2(z)\,dz$$

as $h \to 0$. Condition (2) now follows from (11). To establish (3) we must demonstrate that if $h, k \to 0$ such that $h/k \to 1$, then

$$h^{-1} \operatorname{cov}\{K((x-X_1)/h), K((x-X_1)/k)\} \to 1.$$

But $E[K((x-X_1)/h)] = o(h^{\frac{1}{2}})$, and so in view of the asymptotic formula for σ_h^2 it suffices to prove that

$$h^{-1} \int_{-\infty}^{\infty} [K((x-y)/h) - K((x-y)/k)]^2\,dF(y) \to 0. \tag{13}$$

For any $\varepsilon > 0$ we define

$$I(h,k) = \int_{|x-y|<\varepsilon} [K((x-y)/h) - K((x-y)/k)]^2\,dF(y)$$

as the limit over increasingly fine dissections $x-\varepsilon = y_0 < y_1 < \ldots < y_n = x+\varepsilon$, of

$$\sum_{i=1}^{n-1} [K((x-y_i)/h) - K((x-y_i)/k)]^2 [F(y_i) - F(y_{i-1})].$$

If F satisfies a Lipshitz condition of order one in the ε-neighbourhood of x then each $F(y_i) - F(y_{i-1})$ is dominated by $M(y_i - y_{i-1})$, where M is a fixed constant. Therefore

$$\begin{aligned} I(h,k) &\leqq M \int_{|x-y|<\varepsilon} [K((x-y)/h) - K((x-y)/k)]^2\,dy \\ &\leqq Mh \int_{-\infty}^{\infty} [K(z) - K(zh/k)]^2\,dz. \end{aligned}$$

Since K is of bounded variation then it is continuous almost everywhere (a.e.), and so $K(zh/k) \to K(z)$ a.e. It follows that $I(h,k) = o(h)$ as $h, k \to \infty$. Furthermore,

$$\begin{aligned} &\int_{|x-y|\geqq\varepsilon} [K((x-y)/h) - K((x-y)/k)]^2\,dF(y) \\ &\leqq \varepsilon^{-2} h^2 \int_{-\infty}^{\infty} \{h^{-1}(x-y)[K(h^{-1}(x-y)) - K(k^{-1}(x-y))]\}^2\,dF(y) \\ &= O(h^2), \end{aligned}$$

and so (13) holds.

4. Trigonometric Series Estimators

There is a wide variety of estimators based on trigonometric series or Fourier transforms. We consider only three of them here – the estimator based on the

Fourier series of a density on $(-\pi, \pi)$, that based on the Fejér form of the series, and that based on the cosine series of a density on $(0, \pi)$.

Let $X_1, X_2, \ldots$ be independent random variables whose common distribution has its support confined to $(-\pi, \pi)$. If $x \in (-\pi, \pi)$ and $F'(x) = f(x) \neq 0$, two estimators of $f(x)$ are

$$\hat{f}_{n1}(x) = \hat{f}_{n1}(x; m) = (2\pi)^{-1}\left[1 + 2\sum_{i=1}^{m} (\hat{a}_{ni}\cos ix + \hat{b}_{ni}\sin ix)\right]$$

and

$$\hat{f}_{n2}(x) = (m+1)^{-1}\sum_{i=1}^{m} \hat{f}_{n1}(x; i)$$
$$= (2\pi)^{-1}\left[1 + 2\sum_{i=1}^{m} (1 - i/(m+1))(\hat{a}_{ni}\cos ix + \hat{b}_{ni}\sin ix)\right],$$

where $m = m(n)$ is a sequence of integers tending to infinity and

$$\hat{a}_{ni} = n^{-1}\sum_{j=1}^{n} \cos(iX_j) \quad \text{and} \quad \hat{b}_{ni} = n^{-1}\sum_{j=1}^{n} \sin(iX_j).$$

If the distribution has its support confined to $(0, \pi)$, and $x \in (0, \pi)$, we may use the estimator

$$\hat{f}_{n3}(x) = \pi^{-1}\left(1 + 2\sum_{i=1}^{m} \hat{a}_{ni}\cos ix\right).$$

We shall assume that the common distribution function F has a derivative f in a neighbourhood of x, continuous at x, and that

$$\lim_{\varepsilon \to 0}\limsup_{n \to \infty}\sup_{p} |m(p)/m(n) - 1| = 0, \tag{14}$$

where the inner supremum is taken over integers p with $|p - n| \leq n\varepsilon$.

Theorem 3. *Under the conditions above,*

$$\limsup_{n \to \infty} \pm [\hat{f}_{ni}(x) - E\hat{f}_{ni}(x)](n/2m \log\log n)^{\frac{1}{2}} = [f(x)/\pi]^{\frac{1}{2}} \text{ a.s.}$$

for $i = 1$ *and* 3 *if*

$$m(\log n)^6/n \log\log n \to 0, \tag{15}$$

and

$$\limsup_{n \to \infty} \pm [\hat{f}_{n2}(x) - E\hat{f}_{n2}(x)](n/2m \log\log n)^{\frac{1}{2}} = [f(x)/3\pi]^{\frac{1}{2}} \text{ a.s.}$$

if

$$m(\log n)^4/n \log\log n \to 0. \tag{16}$$

Conditions (14)–(16) would be satisfied in practice, since it is usual to take $m(n) \sim an^b$ for positive numbers a and b with $b < 1$.

Proof. We shall consider only $\hat{f}_{n1}$ and $\hat{f}_{n2}$. In the case of $\hat{f}_{n1}$ we may write

$$\hat{f}_{n1}(x)=n^{-1}\sum_{i=1}^{n} D_m(x-X_i)$$

where

$$D_m(z)=\sin[(2m+1)z/2]/2\pi\sin(z/2)$$

is the Dirichlet kernel. (Note that

$$\hat{f}_{n3}(x)=n^{-1}\sum_{i=1}^{n}[D_m(x-X_i)+D_m(x+X_i)].)$$

Set $b=-a=\pi$ and $K_m(y)=D_m(x-y)$ in Theorem 1. It can be shown that

$$\begin{aligned}4\pi\sin^2(z/2)D'_m(z)=&(2m+1)[\sin(z/2)-(z/2)]\cos[(2m+1)z/2]\\&+[1-\cos(z/2)]\sin[(2m+1)z/2]\\&-(z/2)^2(2m+1)^2\int_0^1 t\sin[(2m+1)tz/2]\,dt,\end{aligned}$$

and so

$$|D'_m(z)|\leqq C\left[m+(2m+1)^2\left|\int_0^1 t\sin[(2m+1)tz/2]\,dt\right|\right]$$

uniformly in $|z|\leqq\pi$. Therefore

$$\int_{-\pi}^{\pi}|D'_m(z)|\,dz\leqq C\left[m+\int_{-\pi}^{\pi}z^{-2}\left|\int_0^{(2m+1)z/2}u\sin u\,du\right|dz\right].$$

But

$$\left|\int_0^v u\sin u\,du\right|\leqq 2|v|\min(1,v^2),$$

and so

$$\int_{-\pi}^{\pi}|D'_m(z)|\,dz=O(m\log m)$$

as $m\to\infty$. It follows that

$$\int_{-\pi}^{\pi}|dK_m(z)|\,dz=O(m\log m). \tag{17}$$

Suppose F has a derivative in the ε-neighbourhood of x. Now,

$$\int_{|x-y|\geqq\varepsilon} D_m^2(x-y)\,dF(y)=O(1)$$

as $m\to\infty$, and so

$$\begin{aligned} I_m &= \int_{-\pi}^{\pi} D_m^2(x-y)\,dF(y) \\ &= \int_{|x-y|<\varepsilon} D_m^2(x-y)\,dF(y) + O(1) \\ &= f(x)\int_{-\pi}^{\pi} D_m^2(x-y)\,dy + \int_{|x-y|<\varepsilon} D_m^2(x-y)[f(y)-f(x)]\,dy + O(1). \end{aligned}$$

It is easily proved that

$$\int_{-\pi}^{\pi} D_m^2(x-y)\,dy = \int_{-\pi}^{\pi} D_m^2(z)\,dz \sim m/\pi,$$

and the continuity of f at x now ensures that $I_m \sim mf(x)/\pi$. It may also be shown that

$$\begin{aligned} \left|\int_{-\pi}^{\pi} D_m(x-y)\,dF(y)\right| &= \left|\int_{|x-y|<\varepsilon} D_m(x-y)\,f(y)\,dy + O(1)\right| \\ &\leqq C\int_{-\pi}^{\pi} |D_m(z)|\,dz = O(\log m). \end{aligned}$$

(See Butzer and Nessel [6, Proposition 1.2.3, p. 42].) Therefore in the notation of Theorem 1,

$$\sigma_m^2 \sim E[K_m^2(X_1)] \sim mf(x)/\pi. \tag{18}$$

Combining this with (17) we see that (2) follows from (15) in the case of $\hat{f}_{n1}$.

If $p \geqq n$ then

$$D_p(z) - D_n(z) = \pi^{-1} \sum_{n+1}^{p} (\cos jz + \sin jz),$$

and so

$$\int_{-\pi}^{\pi} [D_p(z) - D_n(z)]^2\,dz = (2/\pi)(p-n).$$

From this and (18) follows (3).

Finally we consider the case of $\hat{f}_{n2}$, which we may write as

$$\hat{f}_{n2}(x) = n^{-1} \sum_{i=1}^{n} F_m(x - X_i)$$

where

$$F_m(z) = \{\sin[(m+1)z/2]/\sin(z/2)\}^2/2\pi(m+1)$$

is the Fejér kernel. Let $b = -a = \pi$ and $K_m(y) = F_m(x-y)$ in Theorem 1. It can be shown that

$$2\pi(m+1)\sin^3(z/2)F_m'(z)$$
$$=\sin[(m+1)z/2]\{(m+1)[\sin(z/2)-z/2]\cos[(m+1)z/2]$$
$$+(m+1)(z/2)\cos[(m+1)z/2]-\sin[(m+1)z/2]$$
$$+\sin[(m+1)z/2][1-\cos(z/2)]\}.$$

Since $|\theta\cos\theta-\sin\theta|\leqq 2|\theta|\min(1,\theta^2)$ then

$$|F_m'(z)|\leqq C\{1+[(m+1)/2]^2\min[((m+1)z/2)^{-2},1]\}$$

uniformly in $|z|\leqq\pi$. Therefore

$$\int_{-\pi}^{\pi}|F_m'(z)|\,dz=O(m),$$

and consequently

$$\int_{-\pi}^{\pi}|dK_m(z)|=O(m). \qquad (19)$$

As before we may show that

$$\sigma_m^2\sim E[K_m^2(X_1)]\sim f(x)\int_{-\pi}^{\pi}F_m^2(z)\,dz\sim mf(x)/3\pi.$$

Combining this with (19) we see that (2) follows from (16). Condition (3) may be established as in the case of $\hat{f}_{n1}$.

5. Orthogonal Polynomial Estimators

We consider only the case of an estimator based on the Legendre polynomials. Let $X_1, X_2, \ldots$ be independent random variables having a common absolutely continuous distribution whose density f has its support confined to $(-1,1)$. We shall assume that $(1-y^2)^{-\frac{1}{4}}f(y)$ is integrable on $(-1,1)$. Suppose $x\in(-1,1)$, that f is continuous at x and of bounded variation in a neighbourhood of x, and $f(x)\neq 0$.

The orthonormal Legendre system is defined by

$$p_m(z)=[\tfrac{1}{2}(2m+1)]^{\frac{1}{2}}P_m(z), \qquad m\geqq 0,$$

where the functions P_m are the Legendre polynomials. An estimator of $f(x)$ is given by

$$\hat{f}_n(x)=\sum_{i=0}^{m}\hat{a}_{ni}p_i(x),$$

where $m=m(n)$ is a sequence of integers tending to infinity and

$$\hat{a}_{ni}=n^{-1}\sum_{j=1}^{n}p_i(X_j).$$

Assume that (14) holds and

$$m^3(\log n)^4/n \log\log n \to 0. \tag{20}$$

Theorem 4. *Under the conditions above,*

$$\limsup_{n\to\infty} \pm[\hat{f}_n(x)-E\hat{f}_n(x)](n/2m\log\log n)^{\frac{1}{2}} = [f(x)/\pi]^{\frac{1}{2}}(1-x^2)^{-\frac{1}{4}} \text{ a.s.}$$

Proof. The k'th derivative of P_m, $P_m^{[k]}$, is related to the ultraspherical polynomials $P_m^{(\alpha)}$ by the formula

$$P_m^{[k]}(x) = 2^k(\tfrac{1}{2})_k P_{m-k}^{(k+\frac{1}{2})}(x).$$

From this and the result 7.33.6, p. 167 of Szegö [16] we see that

$$|P_m^{[k]}(\cos\theta)| \leqq C\min(m^{2k}, m^{k-\frac{1}{2}}\theta^{-k-\frac{1}{2}}) \tag{21}$$

for $0<\theta<\pi/2$, where C depends only on k.

The Christoffel-Darboux formula ([13, p. 179]) asserts that

$$\begin{aligned} L_m(x,y) &= \tfrac{1}{2}\sum_{i=0}^{m}(2i+1)P_i(x)P_i(y) \\ &= \tfrac{1}{2}(m+1)(y-x)^{-1}[P_m(x)P_{m+1}(y)-P_{m+1}(x)P_m(y)], \end{aligned}$$

and we may write

$$\hat{f}_n(x) = n^{-1}\sum_{i=1}^{n} L_m(x, X_i).$$

Let $b=-a=1$ and $K_m(y)=L_m(x,y)$ in Theorem 1, and define

$$\begin{aligned} M_m(x,y) &= 2(m+1)^{-1}\,\partial L_m(x,y)/\partial y \\ &= (y-x)^{-2}\{P_m(x)[(y-x)P'_{m+1}(y)-P_{m+1}(y)] \\ &\quad -P_{m+1}(x)[(y-x)P'_m(y)-P_m(y)]\}. \end{aligned}$$

From Taylor's theorem we see that

$$P_m(x) = P_m(y)+(x-y)P'_m(y)+(x-y)^2R_m(x,y)$$

where

$$R_m(x,y) = \int_0^1 (1-t)P''_m(tx+(1-t)y)\,dt.$$

Hence

$$M_m(x,y) = P_m(x)R_{m+1}(x,y)-P_{m+1}(x)R_m(x,y). \tag{22}$$

In order to bound $\int_{-1}^{1}|M_m(x,y)|\,dy$ we estimate

$$\begin{aligned} \Delta_m &= \int_{-1}^{1} dy\int_0^1 (1-t)|P''_m(tx+(1-t)y)|\,dt \\ &= (1+x)^{-1}\int_{-1}^{x}(1+z)|P''_m(z)|\,dz + (1-x)^{-1}\int_x^1 (1-z)|P''_m(z)|\,dz. \end{aligned}$$

We now deduce from (21) that $\Delta_m = O(m^{3/2})$. Again from (21), $P_m(x) = O(m^{-\frac{1}{2}})$, and so by (22),

$$\int_{-1}^{1} |M_m(x, y)|\, dy = O(m).$$

Therefore

$$\int_{-1}^{1} |dK_m(y)| = O(m^2). \tag{23}$$

As in the proof of Theorem 3 we may write

$$\begin{aligned} I_m &= \int_{-1}^{1} L_m^2(x, y)\, dF(y) \\ &= f(x) \int_{-1}^{1} L_m^2(x, y)\, dy + \int_{\substack{|x-y| \leq \varepsilon; \\ -1<y<1}} L_m^2(x, y)[f(y) - f(x)]\, dy \\ &\quad + \int_{\substack{|x-y| > \varepsilon; \\ -1<y<1}} L_m^2(x, y) f(y)\, dy + O(1). \end{aligned}$$

If $|x-y| > \varepsilon$ and $-1 < y < 1$ then in view of (21), $|P_m(y)| \leq C m^{-\frac{1}{2}}(1-y^2)^{-\frac{1}{4}}$, and so

$$\begin{aligned} L_m^2(x, y) &\leq C_1 m^2 [P_m^2(x) P_{m+1}^2(y) + P_{m+1}^2(x) P_m^2(y)] \\ &\leq C_2 m [|P_{m+1}(y)| + |P_m(y)|] \leq C_3 m^{\frac{1}{2}} (1-y^2)^{-\frac{1}{4}}, \end{aligned}$$

where C_1, C_2 and C_3 depend only on ε and x. Since $(1-y^2)^{-\frac{1}{4}} f(y)$ is integrable then

$$I_m = f(x) \int_{-1}^{1} L_m^2(x, y)\, dy + \int_{\substack{|x-y| \leq \varepsilon; \\ -1<y<1}} L_m^2(x, y)[f(y) - f(x)]\, dy + O(m^{\frac{1}{2}}).$$

Now,

$$\int_{-1}^{1} L_m^2(x, y)\, dy = \tfrac{1}{2} \sum_{i=0}^{m} (2i+1) P_i^2(x)$$

and

$$P_m^2(\cos\theta) = (m\pi \sin\theta)^{-1} \{1 + \cos[(2m+1)\theta - \pi/2]\} + O(m^{-5/2})$$

as $m \to \infty$ (Laplace's formula; see for example [13, p 208]), and so

$$I_m \sim m f(x)/\pi(1-x^2)^{\frac{1}{2}}.$$

In view of the conditions imposed on f, $\int_{-1}^{1} L_m(x, y)\, dF(y)$ converges as $m \to \infty$ ([13, p. 235]), and so in the notation of Theorem 1,

$$\sigma_m^2 \sim E[K_m^2(X_1)] \sim m f(x)/\pi(1-x^2)^{\frac{1}{2}}.$$

Condition (2) now follows from (20), and (3) may be proved as in Theorem 3.

Acknowledgement. I am grateful to Professor E.J. Hannan and Dr. C.C. Heyde for their comments on an earlier draft of this paper. The referee provided an exceptionally detailed series of comments which led to further improvements.

References

1. Bluez, J., Bosq, D.: Conditions nécessaires et suffisantes de convergence pour une classe d'estimateurs de la densité. C.R. Acad. Sci. Paris **282**, sér. A, 63–66 (1976)
2. Bluez, J., Bosq, D.: Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales. C.R. Acad. Sci. Paris **282**, sér. A. 1023–1026 (1976)
3. Bosq, D.: Sur l'estimation d'une densité multivariée par une série de fonctions orthogonales. C.R. Acad. Sci. Paris **268**, sér. A, 555–557 (1969)
4. Bosq, D.: Estimation non paramétrique de la densité et de ses dérivées. C.R. Acad. Sci. Paris **269**, sér. A, 1010–1012 (1969)
5. Bosq, D.: Complément à deux notes sur l'estimation de la densité et de ses dérivées. C.R. Acad. Sci. Paris **271**, sér. A, 45 (1970)
6. Butzer, P.L., Nessel, R.J.: Fourier Analysis and Approximation, Volume I. New York: Academic Press 1971
7. Deheuvels, P.: Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. C.R. Acad. Sci. Paris **278**, sér. A, 1217–1220 (1974)
8. Devroye, L.P.: On the pointwise and the integral convergence of recursive kernel estimates of probability densities. Utilitas Math. **15**, 113–128 (1979)
9. Devroye, L.P., Wagner, T.J.: The L_1 convergence of kernel density estimates. Ann. Statist. **7**, 1136–1139 (1979)
10. Eicker, F.: A log log-law for double sequences of random variables, I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **16**, 107–133 (1970)
11. Kiefer, J.: Iterated logarithm analogues for sample quantiles when $P_n \downarrow 0$. Proc. 6th Berkeley Sympos. Math. Statist. Probab., Univ. Calif. 1972, **1**, 227–244
12. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent rv's, and the sample df, I. Z. Wahrscheinlichkeitstheorie verw. Gebiete **32**, 111–131 (1975)
13. Sansone, G.: Orthogonal Functions. New York: Interscience 1959
14. Silverman, B.W.: Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. **6**, 177–184 (1978)
15. Singh, R.S.: Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density. Ann. Statist. **5**, 394–399 (1977)
16. Szegö, G.: Orthogonal Polynomials. Amer. Math. Soc. Colloq. Pub. No. 23 (1939)
17. Tomkins, R.J.: On the law of the iterated logarithm for double sequences of random variables. Z. Wahrscheinlichkeitstheorie verw. Gebiete **30**, 303–314 (1974)
18. Wegman, E.J., Davies, H.I.: Remarks on some recursive estimators of a probability density. Ann. Statist. **7**, 316–327 (1979)
19. Winter, B.B.: Rate of strong consistency of two nonparametric density estimators. Ann. Statist. **3**, 759–766 (1975)

Received November 20, 1979; in improved form July 17, 1980