ON THE POINTWISE AND THE INTEGRAL CONVERGENCE OF RECURSIVE KERNEL ESTIMATES OF PROBABILITY DENSITIES

Luc P. Devroye*

ABSTRACT. Let $X_1,\ldots,X_n$ be independent R^d-valued random vectors with a common density f, and let f be estimated by $f_n(x) = n^{-1} \Sigma h_i^{-d} K((x-X_i)/h_i)$, where $\{h_n\}$ is a sequence of positive numbers and K is a bounded density on R^d.

Several results related to the weak or the strong pointwise consistency of f are discussed and derived in the first part of the paper. In the second part, weak conditions on $\{h_n\}$ are given insuring that $\int |f_n(x) - f(x)|dx \overset{n}{\to} 0$ in probability (or with probability one) for *all* densities f.

1. *Introduction.*

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random vectors taking values in R^d with a common unknown density f. To estimate recursively the density for each $x \in R^d$ we consider the Wolverton-Wagner-Yamato estimate

(1) $$f_n(x) = n^{-1} \sum_{i=1}^{n} h_i^{-d} K((x-X_i)/h_i)$$

(see [25-27]), where K is a given bounded probability density and $\{h_n\}$ is a sequence of positive numbers satisfying

(2) $$h_n \overset{n}{\to} 0$$

and

(3) $$n h_n^d \overset{n}{\to} \infty .$$

The estimate f_n is a recursive version of the celebrated Parzen-Rosenblatt kernel estimate [20], [23]

* This research was partially sponsored by AFOSR Grants 72-2371 and 77-3385.

UTILITAS MATHEMATICA Vol. 15 (1979), pp. 113-128.

(4) $$g_n(x) = n^{-1} \sum_{i=1}^{n} h_n^{-d} K((x-X_i)/h_n) .$$

For surveys on nonparametric density estimation, the reader is referred to Wegman [24] and Revesz [22].

It is well known that g_n is weakly pointwise consistent if (2) and (3) hold, if f is continuous at x, and if

(5) $$||x||^d K(x) \to 0 \quad \text{as} \quad ||x|| \to \infty$$

(see [6] and [20]). If, to these conditions, we add

(6) $$\sum_{n=1}^{\infty} e^{-\alpha n h_n^d} < \infty \text{ for all } \alpha > 0 ,$$

then g_n is strongly pointwise consistent as well, that is, $g_n(x) \overset{n}{\to} f(x)$ with probability one (w.p.1) [14]. Limiting himself to bounded continuous f, to bounded Riemann-integrable K and to sequences $\{h_n\}$ satisfying $h_n \overset{n}{\to} 0$ and

(7) $$(n+k)\, h_{n+k}^d \ge c\, n\, h_n^d \quad (\text{some } c > 0, \text{ all } n,k > 0) ,$$

Deheuvels [12] shows that $g_n(x) \overset{n}{\to} f(x)$ w.p.1 for all x if and only if

(8) $$n\, h_n^d / \log \log n \overset{n}{\to} \infty .$$

Condition (8) is implied by

(9) $$n\, h_n^d / \log n \overset{n}{\to} \infty ,$$

which in turn implies (6) but is equivalent to (6) for sequences satisfying (7). However, in the absence of (7), neither (6) nor (9) imply Deheuvels' condition (8); e.g., let $nh_n^d = \sqrt{n}$ except when $n = 2^{k^k}$ for some integer k, in which case we take $nh_n^d = \log \log n$.

In the first part of this paper we discuss and prove similar pointwise results for the Wolverton-Wagner-Yamato estimate (1). For other recent work on recursive density estimation, the reader is referred to the papers of Ahmad and Lin [2], Banon [3], Carroll [7], Davies [8], Davies and Wegman [9], and Rejto and Revesz [21]. The trouble with a recursive estimate like (1) is that the contribution of X_i is weighted by h_i^{-d} and that the estimate is therefore sensitive to sudden decreases in the value of h_n. Recently, Deheuvels [10-12] proposed a particularly elegant estimate

$$\bar{f}_n(x) = \left(\sum_{i=1}^{n} h_i^d \right)^{-1} \sum_{i=1}^{n} K((x-X_i)/h_i)$$

which shares with (1) and (4) that it is a density on R^d and which has the nice property that all X_i are seemingly equally weighted. (At least, due to the boundedness of K, the difference $\bar{f}_n - \bar{f}_{n-1}$ will remain small even if h_n is much smaller than $h_{n-1},\ldots,h_1$.) For $d = 1$, Deheuvels [10-11] shows that $\bar{f}_n$ is strongly pointwise consistent if f and K are bounded densities, if f is continuous at x, and if

(10) $$h_n \overset{n}{\to} 0 \quad \text{and} \quad \sum_{n=1}^{\infty} h_n^d = \infty .$$

Pointwise consistency is usually of limited value to the statistician. Fortunately, it is true that for a sequence of densities f_n that is almost everywhere weakly (strongly) pointwise consistent, the in probability (w.p.1) convergence to 0 of $\int |f_n(x) - f(x)|dx$ follows (Glick [16]); that is, f_n is weakly (strongly) consistent in L_1. In the second part of this report a variety of weak conditions on $\{h_n\}$ are derived insuring the weak (strong) consistency in L_1 of f_n, g_n, and $\bar{f}_n$ for *all* densities f (i.e., f is not required to be bounded or even to be almost everywhere continuous).

2. *Main Results.*

Let us first recall a pointwise consistency theorem for the Parzen-Rosenblatt estimate (see [6], [14], [20]).

THEOREM 1. *Let* $\{h_n\}$ *satisfy* (2), (3), *let* K *be a bounded probability density, and let Condition* A *hold.*

Condition A. One of the following is true:

(i) x *is a continuity point of* f *and* (5) *holds,*

(ii) x *is a Lebesgue point of* f *and* f *is bounded on* R^d,

(iii) x *is a Lebesgue point of* f *and* K *has compact support.*

Under these conditions, $g_n(x) \overset{n}{\to} f(x)$ *in probability. If, in addition,* (6) *holds, then* $g_n(x) \overset{n}{\to} f(x)$ *w.p.*1.

It is worth recalling that x is a Lebesgue point for f if

$$\rho^{-d} \int_{||y-x|| \le \rho} |f(y) - f(x)| dy \to 0 \quad \text{as} \quad \rho \to 0 ,$$

that all continuity points are Lebesgue points, and that almost all points x are Lebesgue points. Using the techniques of proof of Theorem 1, it is shown in the Appendix that f_n converges pointwise to f under essentially the same conditions.

THEOREM 2. $f_n(x) \overset{n}{\to} f(x)$ *in probability if* $\{h_n\}$ *satisfies* (2), (3), *if* K *is a bounded probability density, and if Condition A holds. If, additionally,* (8) *is true, then* $f_n(x) \overset{n}{\to} f(x)$ *w.p.*1.

Proof of Theorem 2. The proof parallels in several respects the proof of Theorem 1 which can be found in [14]. The main tool we need is an inequality due to Bennett [4] for the sum of independent random variables $Y_1, \ldots, Y_n$ with $E\{Y_i\} = 0$, $E\{Y_i^2\} \le \sigma^2$, and $|Y_i| \le b$:

$$P\left\{\left| n^{-1} \sum_{i=1}^{n} Y_i \right| \ge \varepsilon\right\} \le 2 \exp\{-n(\varepsilon/2b)((1+\sigma^2/2b\varepsilon) \log (1+2b\varepsilon/\sigma^2) - 1)\}$$

$$\le 2 \exp\{-n\varepsilon^2/2(\sigma^2+b\varepsilon)\}$$

where the latter inequality follows from the fact that $\log (1+u) > 2u/(2+u)$ for all $u > 0$. Now, let $\varepsilon > 0$ be arbitrary and note that

$$|f_n(x) - f(x)| \le |f_n(x) - E\{f_n(x)\}| + |E\{f_n(x)\} - f(x)| .$$

From Lemma 2 (see Appendix) we know that $E\{Y_n\} \overset{n}{\to} f(x)$ if $Y_n = h_n^{-d} K((x-X_n)/h_n)$. Hence, by the Kronecker lemma [18, p. 238],

$$E\{f_n(x)\} = n^{-1} \sum_{i=1}^{n} E\{Y_i\} \overset{n}{\to} f(x) .$$

Next, if $c = \sup_i E\{Y_i\}$ and $M = \sup_x K(x)$, then $E\{Y_i^2\} \le cMh_i^{-d}$ and $|Y_i - E\{Y_i\}| \le Mh_i^{-d}$. Therefore,

$$P\{|f_n(x) - E\{f_n(x)\}| \geq \varepsilon\} = P\left\{\left| n^{-1}\sum_{i=1}^{n}(Y_i - E\{Y_i\})\right| \geq \varepsilon\right\}$$

$$\leq 2 \exp\{-n\varepsilon^2/2(cM\ell_n + M\varepsilon\ell_n)\},$$

where $\ell_n = \sup_{i \leq n} h_i^{-d}$. The in probability part of Theorem 2 follows from Lemma 1 and (3). A well-known version of the strong law of large numbers (Loeve [18, p. 253]) asserts that if $|Y_n| \leq Ln$ for all n and some $L < \infty$, (which is the case here since the h_n are positive numbers and $nh_n^d/\log\log n \overset{n}{\to} \infty$), then

$$n^{-1}\sum_{i=1}^{n}(Y_i - E\{Y_i\}) \overset{n}{\to} 0 \quad \text{w.p.1}$$

if and only if for every $\varepsilon > 0$

$$\sum_{k=0}^{\infty} P\left\{\left| 2^{-k}\sum_{i=2^k+1}^{2^{k+1}}(Y_i - E\{Y_i\})\right| \geq \varepsilon\right\} < \infty .$$

We thus conclude that $f_n(x) - E\{f_n(x)\} \overset{n}{\to} 0$ w.p.1 if

$$\sum_{k=0}^{\infty}\left\{2 \exp -(2^k\varepsilon^2/(2(cM+M\varepsilon)) \inf_{i \leq 2^k} h_i^d\right\} < \infty$$

which is implied by

$$2^k \inf_{i \leq 2^k} h_i^d/\log k \overset{k}{\to} \infty .$$

This in turn follows from $n \inf_{i \leq n} h_i^d/\log\log n \overset{n}{\to} \infty$ and $nh_n^d/\log\log n \overset{n}{\to} \infty$ (Lemma 1).

Remark 1. (Conditions for weak consistency). An application of Chebyshev's inequality allows a relaxation of condition (3) for weak consistency. We will see that it can be replaced by

(11) $$n^{-2}\sum_{i=1}^{n} h_i^{-d} \overset{n}{\to} 0 .$$

(Indeed, by Toeplitz's lemma [18, p. 238], $nh_n^d \overset{n}{\to} \infty$ entails

$$n^{-2}\sum_{i=1}^{n} h_i^{-d} \le n^{-1}\sum_{i=1}^{n}(ih_i)^{-d} \to 0.)$$

The example in which $h_n^d = n^{-1/2}$ except when $n = 2^k$, in which case $h_n^d = n^{-1}$, is one satisfying (11) but not (3). However, for monotone $\{h_n\}$ both conditions are equivalent in view of

$$(2n)^{-2}\sum_{i=1}^{2n} h_i^{-d} \ge (2n)^{-2}\sum_{i=1}^{2n} h_i^{-d} \ge (2n)^{-2}\, nh_n^{-d} = (4nh_n^d)^{-1} .$$

They are also equivalent whenever (7) holds with $c = 1$. To see this, assume that (11) is true but $nh_n^d \not\to \infty$. In view of (7) this in turn implies that $nh_n^d \le M < \infty$. But then

$$n^{-2}\sum_{i=1}^{n} h_i^{-d} \ge n^{-2}\sum_{i=1}^{n} i/M \ge n^2/(2n^2 M) = 1/2M > 0 ,$$

contradicting (11). We note that the weak consistency part of Theorem 2 under Condition A(i) can also be found in Ahmad and Lin [2].

Remark 2. (Conditions for strong consistency). The strong consistency of f_n was earlier established by Davies [8] under stricter conditions. Deheuvels [12] has shown that under weak additional assumptions on f, K, and $\{h_n\}$, (8) is necessary for the strong pointwise consistency of f_n. In the next theorem we will see that (8) can be replaced by

$$\text{(12)} \qquad \sum_{n=1}^{\infty} n^{-2} h_n^{-d} < \infty$$

without compromising the strong pointwise consistency of f_n. The proof of this is based on Kolmogorov's second moment version of the strong law of large numbers. The sequence $\left\{nh_n^d\,(\log n)/(\log\log n)\right\}$ is one satisfying (8) and (9) but not (12) since $\{1/n(\log n)(\log\log n)\}$ is not summable. On the contrary, the sequence

$$h_n^d = \begin{cases} n^{-1/2}, & n \ne 2^k \\ n^{-1}, & n = 2^k \end{cases}$$

satisfies (12) but not (3), and thus certainly not (6), (8), or (9). Thus the conditions (8) and (12) do not imply each other.

Suppose next that $\{h_n\}$ is sufficiently well-behaved so that Deheuvels' quasi-monotonicity condition (7) holds with $c = 1$. With some work one can show that (7) and (12) imply (9):

$$nh_n^d/\log n \overset{n}{\to} \infty .$$

Thus, for nicely behaved $\{h_n\}$ (e.g., nh_n^d is asymptotically monotone), Theorem 2 is stronger than the said theorem in which (12) replaces (8). (The proof of the latter implication uses the facts that for nondecreasing $\{\alpha_n\}$, $\Sigma\, 1/\alpha_n < \infty$ if and only if $\Sigma\, 2^n/\alpha_{2^n} < \infty$, and that $\Sigma\, 1/\alpha_n < \infty$ implies $n\alpha_n \overset{n}{\to} 0$, to conclude, upon taking $\alpha_n = nh_n^d$, that $\Sigma\, 1/(n\alpha_n) < \infty$ yields $\alpha_n/\log n \overset{n}{\to} \infty$.)

We now state the announced theorem.

THEOREM 3. *If* $\{h_n\}$ *satisfies* (2), (11), *if* K *is a bounded density, and if Condition A holds, then* $E\{(f_n(x) - f(x))^2\} \overset{n}{\to} 0$. *If, additionally,* (12) *holds, then* $f_n(x) \overset{n}{\to} f(x)$ *w.p.1 as well.*

Proof of Theorem 3. We need only show that $|f_n(x) - E\{f_n(x)\}| \overset{n}{\to} 0$ in the quadratic mean or with probability one. By Chebyshev's inequality and (11),

$$P\left\{|f_n(x) - E\{f_n(x)\}| \geq \varepsilon\right\} \leq \varepsilon^{-2}\, E\left\{(f_n(x) - E\{f_n(x)\})^2\right\}$$

$$\leq (n\varepsilon)^{-2} \sum_{i=1}^{n} E\{Y_i^2\} \leq (n\varepsilon)^{-2} \sum_{i=1}^{n} cMh_i^{-d} \overset{n}{\to} 0 .$$

This proves the first part of Theorem 3. By Kolmogorov's condition for the strong law of large numbers (see [18, p. 253]) we know that $f_n(x) - E\{f_n(x)\} \overset{n}{\to} 0$ w.p.1 if

$$\sum_{n=1}^{\infty} E\{Y_n^2\}/n^2 < \infty .$$

Using $E\{Y_n^2\} \leq cMh_n^{-d}$, this condition reduces to (12). Actually, the inequality of Hajek and Renyi [17, pp. 258-260] provides us with an upper bound

$$P\left\{\bigcup_{n=k}^{\infty} \{|f_n(x) - E\{f_n(x)\}| \geq \varepsilon\}\right\} \leq (k\varepsilon)^{-2} \sum_{n=1}^{k} E\{Y_n^2\} + \varepsilon^{-2} \sum_{n=k+1}^{\infty} E\{Y_n^2\}/n^2$$

which tends to 0 as $k \to \infty$ for all $\varepsilon > 0$.

Notice that f_n and g_n are valid densities on R^d. Thus, we can apply a theorem of Glick [16] which states that if $\{f_n\}$ is a sequence of densities converging to a density f in probability (w.p.1) for almost all x with respect to Lebesgue measure, and if the f_n are Borel measurable functions of $x, X_1, \ldots, X_n$, then $\int |f_n(x) - f(x)| dx \overset{n}{\to} 0$ in probability (w.p.1). Thus, the following theorem is a corollary to Theorems 2 and 3.

THEOREM 4. *Let* K *be a bounded probability density and let any one of the following conditions hold:*

(i) f *is almost everywhere continuous and* (5) *holds,*
(ii) f *is bounded,*
(iii) K *has compact support.*

If (2) *and* (3) *hold, then* $\int |g_n(x) - f(x)| dx \overset{n}{\to} 0$ *in probability. If also* (6) *holds, then* $\int |g_n(x) - f(x)| dx \overset{n}{\to} 0$ *w.p.1. If* (2) *and* (11) *hold, then* $\int |f_n(x) - f(x)| dx \overset{n}{\to} 0$ *in probability. If* (2)*, and* (8) *or* (12) *hold, then* $\int |f_n(x) - f(x)| dx \overset{n}{\to} 0$ *w.p.1.*

We emphasize that Theorem 4 holds for *all* densities f if K has compact support, i.e., if for some finite M, $K(x) = 0$ whenever $||x|| > M$. The part of Theorem 4 involving g_n was shown by Devroye and Wagner [15] for kernels K satisfying a condition weaker than (iii), namely

$$\int \sup_{y: ||y|| > ||x||} k(y) \, dx < \infty .$$

Remark 3. (Convergence in L_2 of density estimates). Because of its mathematical attractiveness, several authors consider $\int (f_n(x) - f(x))^2 dx$ as a global measure of the deviation of f_n from f. For instance, Wagner [26] shows that $\int (f_n(x)-f(x))^2 dx \overset{n}{\to} 0$ w.p.1 if K is a bounded density with $\int ||x|| K(x) \, dx < \infty$, if h_n is nonincreasing, if $\Sigma \, h_n/n < \infty$, if (12) holds, and if f satisfies a uniform Lipschitz condition.

A variety of results are available regarding $\int (g_n(x)-f(x))^2 dx$ (see Bickel and Rosenblatt [5], Ahmad [1] and Nadaraya [19]). Most of them are concerned with the rate of convergence to 0 of the mean integrated square error. It is known that if K is a bounded density satisfying (5), if (2)-(3) hold, and if f^2 is integrable and f is almost everywhere continuous, then $E\{\int (g_n(x) - f(x))^2 dx\} \overset{n}{\to} 0$ [13]. If in addition

$$nh_n^{2d}/\log n \overset{n}{\to} \infty ,$$

Nadaraya [19] shows that $\int (g_n(x) - f(x))^2 dx \overset{n}{\to} 0$ w.p.1. (The latter condition is not necessary. Nadaraya's argument remains valid if we replace it by (9) and use Bennett's inequality [4] in the proper places.) In any case, all these results are only applicable if f^2 is integrable. Since f is unknown in advance this condition cannot be checked. Theorem 4 shows that the situation is not all that bad since the integrated L_1 error converges to 0 w.p.1 for *all* densities f (under mild conditions on the sequence $\{h_n\}$ which we can pick anyway) and all the estimates discussed in this paper. This result is not surprising since L_1, not L_2, is the natural space in which the properties of densities should be studied. It is comforting to know that the requirement that f be almost everywhere continuous can be dropped altogether.

3. *A Note on the Deheuvels Estimator.*

In this section we complement some results of Deheuvels [10-12].

THEOREM 5. *If* $\{h_n\}$ *satisfies* (10), *if* K *is a bounded probability density and if Condition A holds, then* $\bar{f}_n(x) \overset{n}{\to} f(x)$ *w.p.1.*

Proof of Theorem 5. Notice first that

$$E\{\bar{f}_n(x)\} = \left(\sum_{i=1}^{n} h_i^d\right)^{-1} \sum_{i=1}^{n} h_i^d E\{Y_i\} ,$$

where Y_i is defined as in the proof of Theorem 2. Using Lemma 2 ($E\{Y_n\} \overset{n}{\to} f(x)$) and Toeplitz's lemma [18, p. 238] we see that $\Sigma h_n^d = \infty$ implies that $E\{\bar{f}_n(x)\} \overset{n}{\to} f(x)$. Thus we need only show that $\bar{f}_n(x) - E\{\bar{f}_n(x)\} \overset{n}{\to} 0$ w.p.1. By Loeve's criterion for the strong law of large numbers [18, p. 253] this follows if

$$\sum_{n=1}^{\infty} E\{K^2((x-X_n)/h_n)\} \Big/ \left(\sum_{i=1}^{n} h_i^d\right)^2 < \infty .$$

But we know that $E\{K^2((x-X_n)/h_n)\} \le Mh_n^d E\{Y_n\} \le Mch_n^d$. Theorem 5 now follows since $\sum_{n=1}^{\infty} h_n^d = \infty$ implies

$$\sum_{n=1}^{\infty} h_n^d \Big/ \left(\sum_{i=1}^{n} h_i^d\right)^2 < \infty .$$

To see this, assume that $h_1 > 0$ and notice that

$$\sum_{n=1}^{\infty} h_n^d \Big/ \left(\sum_{i=1}^{n} h_i^d\right)^2 \le 1/h_1^d + \sum_{n=2}^{\infty} h_n^d \Big/ \left(\left(\sum_{i=1}^{n} h_i^d\right)\left(\sum_{i=1}^{n-1} h_i^d\right)\right)$$

$$\le \sum_{n=2}^{\infty} \left(1 \Big/ \sum_{i=1}^{n-1} h_i^d - 1 \Big/ \sum_{i=1}^{n} h_i^d\right) + 1/h_1^d$$

$$= 2/h_1^d .$$

THEOREM 6. *If* $\{h_n\}$ *satisfies* (10), *if* K *is a bounded probability density and if any one of the following conditions hold,*

- *(i)* f *is almost everywhere continuous and* (5) *is satisfied,*
- *(ii)* f *is bounded,*
- *(iii)* K *has compact support,*

then $\int |\bar{f}_n(x) - f(x)| dx \overset{n}{\to} 0$ *w.p.1.*

One should notice that (10) is weaker than (11) since

$$\sum_{i=1}^{n} h_i^d \ge n^2 \left(\sum_{i=1}^{n} h_i^{-d}\right)^{-1} \overset{n}{\to} 0 .$$

Here we used the following inequality valid for all sequences $\{h_n\}$:

$$\left(n^{-1}\sum_{i=1}^{n} h_i^{-d}\right)\left(n^{-1}\sum_{i=1}^{n} h_i^{d}\right) \ge 1 .$$

The conditions (10) cannot be improved upon in Theorem 5 due to a result by Deheuvels [12] who shows that under some additional restrictions on f, K, and $\{h_n\}$, in order for $f_n(x) \overset{n}{\to} f(x)$ w.p.1, it is necessary that (10) hold.

Proof of Theorem 6. Theorem 6 follows from Theorem 5, Glick's result [16], and the fact that almost all x are Lebesgue points for f.

APPENDIX

LEMMA 1. *If* $a_n, b_n \geq 0$, $a_n \uparrow \infty$, *then* $a_n/b_n \overset{n}{\to} \infty$ *if and only if* $a_n/(\sup_{i\leq n} b_i) \overset{n}{\to} \infty$.

Proof. Notice that

$$a_n/b_n \geq a_n/(\sup_{i\leq n} b_i) \geq \min(\inf_{i>N} a_i/b_i, a_n/(\sup_{i\leq N} b_i)) .$$

Lemma 1 follows by first picking N large enough and then letting n grow unbounded.

COROLLARY. *The conditions* $nh_n^d \overset{n}{\to} \infty$ *and* $n \inf_{i\leq n} h_i^d \overset{n}{\to} \infty$ *are equivalent. Furthermore, the conditions* $nh_n^d/\log\log n \overset{n}{\to} \infty$ *and* $n \inf_{i\leq n} h_i^d/\log\log n \overset{n}{\to} \infty$ *are equivalent.*

LEMMA 2. *Let* K *be a bounded probability density, let Condition A hold, and let* $h_n \overset{n}{\to} 0$. *Then*

$$E\left\{h_n^{-d} K((x-X_1)/h_n)\right\} \overset{n}{\to} f(x) .$$

Proof. We need only show that

$$\left|\int h_n^{-d} K((x-y)/h_n) f(y)dy - f(x)\right| \leq \int h_n^{-d} K((x-y)/h_n)\,|f(y)-f(x)|\,dy \overset{n}{\to} 0 . \tag{13}$$

Let $S(\alpha,\rho)$ be a closed sphere in R^d centered at α with radius ρ and let $(\cdot)^c$ denote the complement of a set. We can upper bound (13) by

$$(\rho h_n)^{-d} \int_{S(x,\rho h_n)} M|f(y)-f(x)|dy + \min(\sup_{S^c(x,\rho)} K(y); (\sup f(y))) \int_{S^c(x,\rho)} K(y)dy ,$$

where $M = \sup K(y)$ and ρ is chosen. The last term can be made small by choice of ρ if either K has compact support or f is bounded. The first term is small for large n if $h_n \overset{n}{\to} 0$ and x is a Lebesgue point of f. Expression (13) can also be upper bounded by

$$\sup_{S(x,\rho)} |f(y) - f(x)| + f(x) \int_{S^c(x,\rho/h_n)} K(y)dy + \rho^{-d} \sup_{S^c(x,\rho/h_n)} ||y||^d K(y) .$$

If x is a continuity point of f, then we can make the first term small by choice of ρ. The last two terms are small for large n if (5) holds, if K is integrable, and if $h_n \overset{n}{\to} 0$.

REFERENCES

[1] I. A. Ahmad, *Strong convergence of a quadratic measure of the deviation of density function estimates*, Rep. Stat. Applic. Res. JUSE *23* (1976), 1-5.

[2] I. A. Ahmad and P. Lin, *Nonparametric sequential estimation of a multiple regression function*, Bull. Math. Stat. *17* (1976), 63-75.

[3] G. Banon, *Sur un estimateur nonparamétrique de la densité de probabilité*, Revue de Statistique Appliquée *24* (1976), 61-73.

[4] G. Bennett, *Probability inequalities for the sum of independent random variables*, J. Amer. Stat. Assoc. *57* (1962), 33-45.

[5] P. J. Bickel and M. Rosenblatt, *On some global measures of the deviation of density function estimate*, Ann. Stat. *1* (1973), 1079-1095.

[6] T. Cacoullos, *Estimation of a multivariate density*, Ann. Inst. Stat. Math. *18* (1965), 179-190.

[7] R. J. Carroll, *On sequential density estimation*, Z. für Wahrscheinlichkeitstheorie und verwandte Gebiete *36* (1976), 136-151.

[8] H. I. Davies, *Strong consistency of a sequential estimator of a probability density function*, Bull. Math. Stat. *15* (1973), 49-53.

[9] H. I. Davies and E. J. Wegman, *Sequential nonparametric density estimation*, IEEE Trans. on Inf. Thy. *21* (1975), 619-628.

[10] P. Deheuvels, *Sur une famille d'estimateurs de la densité d'une variable aléatoire*, C. R. de l'Acad. des Sci. de Paris *276* (1973), 1013-1015.

[11] P. Deheuvels, *Sur l'estimation séquentielle de la densité*, C. R. de l'Acad. des Sci. de Paris *276* (1973), 1119-1121.

[12] P. Deheuvels, *Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité*, C. R. de l'Acad. des Sci. de Paris *278* (1974), 1217-1220.

[13] L. P. Devroye and T. J. Wagner, *Nonparametric discrimination and density estimation*, Technical Report #183, The University of Texas at Austin, 1976.

[14] L. P. Devroye and T. J. Wagner, *The strong uniform consistency of kernel density estimates,* to appear in Multivariate Analysis *V*, P. R. Krishnaiah, Ed., (1978).

[15] L. P. Devroye and T. J. Wagner, *The* L_1 *convergence of kernel density estimates,* to appear in Annals of Statistics, (1978).

[16] N. Glick, *Consistency conditions for probability estimators and integrals of density estimators,* Util. Math. *6* (1974), 61-74.

[17] B. V. Gnedenko, *The Theory of Probability,* Chelsea, New York, 1967.

[18] M. Loeve, *Probability Theory,* Van Nostrand, Princeton, New Jersey, 1963.

[19] E. A. Nadaraya, *On the convergence in the* L_2 *norm of probability density estimates,* Theory of Probability and its Applications *18* (1973), 808-817.

[20] E. Parzen, *On the estimation of a probability density function and the mode,* Ann. Math. Stat. *33* (1962), 1065-1076.

[21] L. Rejto and P. Revesz, *Density estimation and pattern classification,* Problems of Control and Information Theory *2* (1973), 67-80.

[22] P. Revesz, *On empirical density function,* Periodica Mathematica Hungarica *2* (1972), 85-110.

[23] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function,* Ann. Math. Stat. *27* (1957), 832-837.

[24] E. J. Wegman, *Nonparametric probability density estimation: I. A summary of available methods,* Technometrics *14* (1972), 533-546.

[25] C. T. Wolverton and T. J. Wagner, *Asymptotically optimal discriminant functions for pattern classification,* IEEE Trans. on Inf. Thy. *15* (1969), 258-265.

[26] C. T. Wolverton and T. J. Wagner, *Recursive estimates of probability densities,* IEEE Transactions on Systems Science and Cybernetics *5* (1969), 307.

[27] H. Yamato, *Sequential estimation of a continuous probability density function and the mode*, Bull. Math. Stat. *14* (1971), 1-12.

School of Computer Science
McGill University
Montreal H3A 2K6

Received March 25, 1977; revised September 26, 1977.