*The Canadian Journal of Statistics*
Vol. 14, No. 3, 1986, Pages 211–219
*La Revue Canadienne de Statistique*

# The strong uniform convergence of multivariate variable kernel estimates*

Luc DEVROYE and Clark S. PENROD

*McGill University* and *The University of Texas at Austin*

*Key words and phrases*: Density estimation, consistency, strong convergence, kernel estimate, Breiman's estimate.
*AMS* 1980 *subject classifications*: Primary 60F15; secondary 62G99, 62H99.

## ABSTRACT

We show that $\sup_x |f_n(x) - f(x)| \to 0$ completely as $n \to \infty$, where $f$ is a uniformly continuous density on $\mathbb{R}^d$, $X_1, \ldots, X_n$ are independent random vectors with common density $f$, and $f_n$ is the variable kernel estimate

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{H_{ni}^d} K\left(\frac{X_i - x}{H_{ni}}\right).$$

Here $H_{ni}$ is the distance between $X_i$ and its $k$th nearest neighbour, $K$ is a given density satisfying some regularity conditions, and $k$ is a sequence of integers with the property that $k/n \to 0$, $k/\log n \to \infty$ as $n \to \infty$.

## RÉSUMÉ

Nous démontrons que $\sup_x |f_n(x) - f(x)| \to 0$ complètement lorsque $n \to \infty$, où $f$ représente une fonction de densité uniformément continue sur $\mathbb{R}^d$, $X_1, \ldots, X_n$ sont des vecteurs aléatoires indépendants ayant $f$ pour densité commune, et où $f_n$ dénote l'estimateur du noyau variable, à savoir

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{H_{ni}^d} K\left(\frac{X_i - x}{H_{ni}}\right).$$

Ici, $H_{ni}$ représente la distance entre $X_i$ et son $k$-ième voisin le plus proche, $K$ est une fonction de densité donnée obéissant à quelques conditions de régularité et $k$ est une suite d'entiers telle que $k/n \to 0$ et $k/\log n \to \infty$ quand $n \to \infty$.

## 1. INTRODUCTION

We would often like to estimate a density $f$ on $\mathbb{R}^d$ from $X_1, \ldots, X_n$, a sample of independent identically distributed random vectors (with density $f$) by a function $f_n$ which itself is a density. Estimates which satisfy this requirement include the *histogram estimate* and the *kernel estimate*, among others [see Wertz (1978), Wertz and Schneider (1979), or Tapia and Thompson (1978) and Devroye and Gyorfi (1985) for references on density estimation].

Recent work in the area of density estimation has been in the direction of improved small-sample performance and automatization (i.e., choice of the parameters as a function

* Research of both authors was supported by the Office of Naval Research, Contract N00014-81-K-0145.

of the data). In particular, the choice of the smoothing parameter in the kernel estimate has drawn a great deal of attention (Duin 1976; Deheuvels 1977a, b; Scott, Tapia, and Thompson 1977; Silverman 1978; Scott and Factor 1981; Nadaraya 1974; Devroye and Wagner 1980; Chow, Geman, and Wu 1982; Rudemo 1982; Schuster and Gregory 1981). The theoretical analysis of the kernel estimate indicates that in addition to data dependence one should also let the smoothing parameter depend upon $x$, the point at which $f$ is estimated. In general, however, this leads to estimates that are not densities. For example, for the kernel estimate

$$f_n(x) = \frac{1}{nH_n^d} \sum_{i=1}^{n} K\left(\frac{X_i - x}{H_n}\right),$$

where $K$ is a given density and $H_n$ is the smoothing parameter, Moore and Yackel (1977) and Mack and Rosenblatt (1979) analyze the bias and variance of $f_n$ when $H_n = H_n(x)$ is the distance from $x$ to its $k$th nearest neighbour. Note that this estimate generalizes the $k$-nearest-neighbour estimate of Loftsgaarden and Quesenberry (1965):

$$f_n^{\dagger}(x) = \frac{k/n}{cH_n^d(x)},$$

where $c$ is the volume of the unit sphere in $\mathbb{R}^d$. These estimates have the appealing property of local smoothing, which will be advantageous when the unknown density $f$ is not smooth. However, they consistently overestimate $f$ in the tail regions, resulting in density estimates which have infinite integrals.

Several techniques have been suggested for transforming the $k$-nearest-neighbour estimate (with its infinite integral but appealing local smoothing) into an estimate that integrates to one. Most of these are of the following type (which we shall call the *variable-kernel estimate*):

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{H_{ni}^d} K\left(\frac{X_i - x}{H_{ni}}\right), \tag{1}$$

where $H_{ni}$ is the distance from $X_i$ to its $k$th nearest neighbour among $X_j$, $j \neq i$, and $k = k_n$ is a positive integer. The estimate (1) is a simplified version of an estimate of Breiman, Meisel, and Purcell (1977), who also propose to choose $k$ as a function of the data, although they report that the choice of $k$ over a wide range has surprisingly little effect on the performance of $f_n$. For $d = 1$, (1) is similar to an estimate suggested by Wagner (1975). Note that this estimate provides for local smoothing; however, it does yield a density. Wertz mentions that good results with estimates similar to (1) were obtained by Professor N. Victor of Giesen (Wertz 1978, p. 59). For experimental comparisons between (1) and other estimates, see Raatgever and Duin (1978) and Habbema, Hermans, and Remme (1978).

Our results do not apply to the case in which $k$ is a function of the data. In this respect, the estimate studied here cannot be called adaptive, or automatic.

THEOREM 1. *If*

$$K(x) = I_{\|x\| \le 1/c}, \tag{2}$$

*where $I$ is the indicator function, $c$ is a normalization constant,*

$$\lim_{n\to\infty} \frac{k}{n} = 0, \tag{3}$$

$$\lim_{n\to\infty} \frac{k}{\log n} = \infty, \tag{4}$$

*and f is uniformly continuous, then the variable kernel estimate defined by* (1) *and* (2) *satisfies the following; for all* $\epsilon > 0$ *there exists* $\delta > 0$ *and integer* $n_0$ *such that*

$$P(\sup_x |f_n(x) - f(x)| > \epsilon) \leq e^{-\delta k}, \qquad n \geq n_0. \tag{5}$$

REMARK 1. The upper bound in (5) is summable in $n$. Thus Theorem 1 implies that $f_n$ is completely (and thus strongly) uniformly consistent.

REMARK 2. Since $f_n$ is a density, we note that (5) implies that

$$\int |f_n - f| \to 0 \quad \text{a.s.} \qquad \text{as} \quad n \to \infty,$$

although the conditions of the theorem can be relaxed for the $L_1$ convergence. In fact, if $K$ decreases along rays, $k \to \infty$, and $k/n \to 0$, then $\int |f_n - f| \to 0$ in probability for all $f$ (Devroye 1985).

REMARK 3. In view of the general results of Abou-Jaoude (1974), the condition that $f$ is uniformly continuous comes close to being necessary. For example, Abou-Jaoude has shown that no density estimate can be weakly uniformly consistent for all continuous densities on $\mathbb{R}^d$.

The theorem can be extended to other kernels. One possible extension requiring little extra work is given in Theorem 2.

THEOREM 2. *When f is uniformly continuous, and* (3) *and* (4) *hold, then the estimate* (1) *satisfies the conclusion of Theorem 1 (i.e.,* (5)*) for all kernels K of the following type:* $K(x) = L(\|x\|)$ *for some function L vanishing outside* $[0,1]$, *bounded, nonincreasing, and satisfying*

$$\int_0^1 c d u^{d-1} L(u)\, du = 1$$

*(this is equivalent to asking that* $\int K = 1$.*)*

## 2. PROOFS

LEMMA 1. *When* (3) *and* (4) *hold, and f is uniformly continuous, then* $f_n^\dagger$ *satisfies* (5).

*Proof.* Lemma 1 follows after a careful analysis of the results in Devroye and Wagner (1977). Q.E.D.

LEMMA 2. *There exists a constant A only depending upon d such that*

$$\sum_{i=1}^{n} I_{\|X_i - x\| \leq H_{ni}} \leq Ak.$$

*Proof.* Find $A$ cones centered at $x$ such that if $y$ and $z$ belong to the same cone, then $\|x - y\| \leq \|x - z\|$ implies $\|y - z\| < \|x - z\|$. These cones may be overlapping. Among the $X_i$'s in the same cone, we have $\|X_i - x\| \leq H_{ni}$ at most $k$ times. Q.E.D.

LEMMA 3. *Let* $M > 0$ *be a constant, and let T be the set* $\{x : \|x\| \leq M\} \cap \text{support}(X_1)$. *Then* $\sup_T H_n(x) \to 0$ *exponentially when* (3) *holds, i.e. for all* $\epsilon > 0$ *there exist* $\delta$, $n_0 > 0$ *such that* $P(\sup_T H_n(x) > \epsilon) \leq e^{-\delta n}$, $n \geq n_0$. *This property remains valid for all distributions of* $X_1$.

*Proof.* For fixed $\epsilon > 0$ find a finite number of points $x_1, \ldots, x_N$ in $T$ such that $T$ is covered by the union of $S_{x_i \epsilon/2}$, $1 \leq i \leq N$, where $S_{xr}$ denotes the closed sphere of radius $r$ centered at $\boldsymbol{x}$. Let $p = \inf_i \mu(S_{x_i\epsilon/4})$, where $\mu$ is the probability measure of $X_1$. By the definition of the support of a random variable, we have $p > 0$. If $Z$ is a binomial $(n, p)$ random variable, we have by Hoeffding's inequality (Hoeffding 1963)

$$P\left(\sup_T H_n(\boldsymbol{x}) > \epsilon\right) \leq P\left(\bigcup_{i=1}^{N} \left\{\sum_{j=1}^{n} I_{X_j \in S_{x_i\epsilon/4}} < k\right\}\right) \leq NP(Z < k)$$

$$\leq NP\left(Z - np < -\frac{np}{2}\right)$$

$$\leq N \exp\left\{-2n\left(\frac{p}{2}\right)^2\right\} = n \exp\left(-\frac{np^2}{2}\right)$$

whenever $k/n < p/2$. This concludes the proof of Lemma 3. Q.E.D.

LEMMA 4. *Let $V_r$ be the class of all closed spheres in $\mathbb{R}^d$ with diameter not greater than $r$. Let $\mu$ be a probability measure on the Borel sets of $\mathbb{R}^d$ such that*

$$\sup_{V_r} \mu(A) \leq b \leq \frac{1}{4}.$$

*If $\mu_n$ is the empirical measure for $X_1, \ldots, X_n$, a sample of independently and identically distributed random vectors with common probability measure $\mu$, then*

$$P(\sup_{V_r} |\mu_n(A) - \mu(A)| > \epsilon) \leq 8(2n)^{d+1} e^{-n\epsilon^2/(64b+4\epsilon)} + 8ne^{-nb/10}$$

*for all $n$ such that $n \geq \max(1/b, 8b/\epsilon^2)$.*

*Proof.* We refer to Devroye and Wagner (1980, pp. 64–65): use (2.3) and (2.4) in this reference, and note that, in the notation of the reference, $s(\mathcal{A}, 2n) \leq 2(2n)^{d+1}$. Q.E.D.

LEMMA 5. *Let $\epsilon > 0$ be arbitrary and let $B_\epsilon = \{\boldsymbol{x} : f(\boldsymbol{x}) \geq \epsilon\}$. Then* (3) *implies*

$$\sup_{\substack{y : \|y - x\| \leq H_n(y) \\ x \in B_\epsilon}} H_n(\boldsymbol{y}) \to 0$$

*exponentially as $n \to \infty$, for all uniformly continuous densities $f$.*

*Proof.* Let $\delta > 0$ be so small that $|f(\boldsymbol{y}) - f(\boldsymbol{x})| \leq \epsilon/2$ whenever $\|\boldsymbol{x} - \boldsymbol{y}\| \leq \delta$. For the supremum in Lemma 5 we have for $n > \delta$

$$[\sup H_n(\boldsymbol{y}) > \eta] = [\sup H_n(\boldsymbol{y}) > \delta] \cup [\sup H_n(\boldsymbol{y}) \in (\eta, \delta]].$$

Now, we introduce the notion of a tangent sphere $T_{x\delta}(z)$: it is the closed sphere with center *on* $S_{x\delta}$ and radius $\delta$ having the property that $\boldsymbol{z} - \boldsymbol{x}$, the centre, and $\boldsymbol{x}$ are collinear. Thus, $\boldsymbol{x} - \boldsymbol{x}$ gives the "direction" of the sphere. Let $\boldsymbol{x}, \boldsymbol{y}$ be points with $\|\boldsymbol{x} - \boldsymbol{y}\| \leq H_n(\boldsymbol{y})$, $H_n(\boldsymbol{y}) > \delta$. Then $T_{x\delta/2}(y)$ has less than $k$ points, because it is entirely contained in $S_{yH_n(y)}$. But the center of $T_{x\delta/2}(y)$ belongs to $B_{\epsilon/2}$, and thus

$$P\left(\sup_{\substack{\|x - y\| \leq H_n(y) \\ x \in B_\epsilon}} H_n(\boldsymbol{y}) > \delta\right) \leq P\left(\sup_{B_{\epsilon/2}} H_n(\boldsymbol{x}) > \frac{\delta}{2}\right).$$

Also, omitting the object of the supremum when $\|\boldsymbol{x} - \boldsymbol{y}\| \leq H_n(\boldsymbol{y})$, $\boldsymbol{x} \in B_\epsilon$ is meant, we have

$$P(\sup H_n(y) \in (\eta, \delta]) \le P\left(\sup_{B_{\epsilon/2}} H_n(x) > \eta\right),$$

Because $x \in B_\epsilon$, $\|x - y\| \le H_n(y) \le \delta$ implies that $y \in B_{\epsilon/2}$. Thus, for all $\eta > 0$

$$P(\sup H_n(y) > \eta) \le 2P\left(\sup_{B_{\epsilon/2}} H_n(x) > \min\left(\eta, \frac{\delta}{2}\right)\right),$$

and this tends to 0 exponentially by Lemma 3. Q.E.D.

*Proof of Theorem 1.* We note that

$$|f_n(x) - f(x)| \le \sum_{i=1}^{n} U_{ni}(x),$$

where

$$U_{n1}(x) = \frac{1}{k}\sum_{i=1}^{n} \left|\frac{k}{n}(cH_{ni}^d)^{-1} - f(X_i)\right| I_{\|X_i - x\| \le H_{ni}},$$

$$U_{n2}(x) = \frac{1}{k}\sum_{i=1}^{n} \left|f(X_i) - f(x)\right| I_{\|X_i - x\| \le H_{ni}},$$

and

$$U_{n3}(x) = f(x)\left|\frac{1}{k}\sum_{i=1}^{n} I_{\|X_i - x\| \le H_{ni}} - 1\right|.$$

*Part 1.* Since

$$\frac{k}{n}(cH_{ni}^d)^{-1} = \frac{k}{k+1}\,\frac{k+1}{k}\{cH_n^d(X_i)\}^{-1} = \frac{k}{k+1} f_n^\dagger(X_i),$$

where $f_n^\dagger$, $H_n^d(\cdot)$ are defined as in the introduction with $k$ replaced by $k + 1$, we have

$$\sup_x U_{n1}(x) \le A \sup_x \left|\frac{k}{k+1} f_n^\dagger(x) - f(x)\right| \qquad \text{(by Lemma 2)}$$

$$\le A \sup_x |f_n^\dagger(x) - f(x)| + \frac{A}{k}\sup_x f(x) \to 0.$$

Now, the last term is o (1), and the first term satisfies (5) by Lemma 1.

*Part 2.* Let $\delta > 0$ be so small that $|f(y) - f(x)| < \epsilon$ whenever $\|x - y\| \le \delta$. Let $D_n$ be the event $\{\sup_{B_\epsilon} H_n(x) > \delta\}$, where $B_\epsilon$ is as in Lemma 5. Let $(\cdot)^c$ denote the complement of a set or event. Now,

$$\sup_x U_{n2}(x) \le \sup_{B_\epsilon} U_{n2}(x) I_{D_n} + \sup_{B_\epsilon} U_{n2}(x) I_{D_n^c} + \sup_{B_\epsilon^c} U_{n2}(x)$$
$$= \text{I} + \text{II} + \text{III}. \tag{6}$$

By Lemmas 2 and 3, we can conclude that $I \le A \sup_x f(x) I_{D_n} \to 0$ exponentially. The sum in II can be split into a sum over all $i$ with $H_{ni} < \delta$ and over all remaining $i$. The first subsum does not exceed $\epsilon A$ by Lemma 2, while the second subsum is not greater than

$$\frac{n}{k}\sup_x f(x) I_{D_n^*}, \qquad \text{where} \quad D_n^* = \left\{\sup_{\substack{\|x - y\| \le H_n(y) \\ x \in B_\epsilon}} H_n(y) \ge \delta\right\}.$$

But this term tends to 0 exponentially in view of Lemma 5. To treat III, we define the event

$$E_n = \left\{ \sup_{x : H_n(x) > \delta} f(x) \geq 2\epsilon \right\} = \left\{ \sup_{B_{2\epsilon}} H_n(x) > \delta \right\}.$$

Thus,

$$\text{III} \leq A \sup_x f(x)\, I_{E_n} + \frac{1}{k} \sum_{i=1}^{n} f(X_i) I_{\|X_i - x\| \leq H_{ni}} I_{E_n^c} = \text{IV} + \text{V}. \tag{7}$$

By Lemma 3, IV tends to 0 at an exponential rate. Next, on $E_n^c$ we either have $H_{ni} \leq \delta$ [in which case $f(X_i) \leq f(x) + \epsilon \leq 2\epsilon$, since $x \notin B_\epsilon$], or $H_{ni} > \delta$ [in which case $f(X_i) \leq 2\epsilon$ by definition of $E_n^c$]. Thus, applying Lemma 2 again, we see that $\text{V} \leq 2\epsilon A$. Because $\epsilon$ was arbitrary, we have shown that $\sup_x U_{n2}(x) \to 0$ exponentially.

*Part 3.* Consider $\sup_x U_{n3}(x)$. Let $\epsilon > 0$ be arbitrary, and let $M, M^*$ be constants greater than one. Let $\delta > 0$ be such that $\|x - y\| \leq \delta$ implies $|f(x) - f(y)| < \epsilon/M^*$. Let $B = B_\epsilon$, and assume that $M$ is so large that $B \subseteq T = S_{0M} \cap \text{support}(X_1)$. Let $D_n$ be the event

$$\left\{ \sup_T H_n(x) > \delta \right\} \cup \left\{ \sup_{\substack{\|x - y\| \leq H_n(y) \\ x \in B}} H_n(y) \geq \delta \right\}.$$

Let $f_n^\dagger$ be the Loftsgaarden–Quesenberry estimate with $k + 1$ instead of $k$ (see introduction), and let $E_n$ be the event

$$\left\{ \sup_x \left| \frac{k}{k+1} f_n^\dagger(x) - f(x) \right| \geq \frac{\epsilon}{M^*} \right\}.$$

It is clear that $\sup_{B^c} U_{n3}(x) \leq A\epsilon$. Now, let $G_1, \ldots, G_N$ be a cover of $T$ consisting of nonoverlapping rectangles of diameter not exceeding $\delta$. Let $g_i = \sup_{G_i} f(x)$. On $D_n^c E_n^c$, we have, for all $x \in BG_i$ and for all $j$ with $\|X_j - x\| \leq H_{nj}$,

$$\begin{aligned} \left| \frac{k}{n} (cH_{nj}^d)^{-1} - g_i \right| &\leq \left| \frac{k}{n} (cH_{nj}^d)^{-1} - f(X_j) \right| + |f(X_j) - f(x)| + |f(x) - g_i| \\ &\leq 3\epsilon/M^*. \end{aligned}$$

Thus,

$$\frac{\sup_B U_{n3}(x)}{\sup_x f(x)} \leq AI_{D_n} + AI_{E_n} + I_{D_n^c E_n^c} \max_i \sup_{BG_i : g_i > 0} V_{ni}(x), \tag{8}$$

where

$$V_{ni}(x) = \max \left( \frac{1}{k} \sum_{j=1}^{n} I_{\|X_j - x\| \leq r_-} - 1,\ 1 - \frac{1}{k} \sum_{j=1}^{n} I_{\|X_j - x\| \leq r_+} \right)$$

and

$$r_+ = \left( \frac{k}{\text{cn}(g_i + 3\epsilon/M^*)} \right)^{1/d}, \qquad r_- = \left( \frac{k}{\text{cn}(g_i - 3\epsilon/M^*)} \right)^{1/d}.$$

This is well defined when $\inf_{i : g_i > 0} g_i > 3\epsilon/M^*$. The terms $AI_{D_n}$ and $AI_{E_n}$ in (8) satisfy (5) by Lemmas 1, 3, and 5. For the last term in (8) we consider without loss of generality only the case $i = 1$, $g_1 > 0$. Choose $M^*$ large enough so that $g_1 > 3\epsilon/M^*$. Let $\mu$ and $\mu_n$ be as in Lemma 4. Then,

$$V_{n1}(x) = \max\left(\frac{1}{k}\, n\mu_n(S_{xr_-}) - 1,\ 1 - \frac{1}{k}\, n\mu_n(S_{xr_+})\right)$$

$$\le \max\left(\frac{n}{k}\, \mu(S_{xr_-}) - 1,\ 1 - \frac{n}{k}\, \mu(S_{xr_+})\right)$$

$$+ \sup_{r \le r_-} \frac{n}{k}\left|\mu_n(S_{xr}) - \mu(S_{xr})\right|. \tag{9}$$

For all $n$ large enough, we have $r_+ \le r_- < \delta$. For such $n$,

$$\left(g_1 - \frac{2\epsilon}{M^*}\right) cr^d \le \left(f(x) - \frac{\epsilon}{M^*}\right) cr^d \le \mu(S_{xr}) \le \left(f(x) + \frac{\epsilon}{M^*}\right) cr^d$$

$$\le \left(g_1 + \frac{\epsilon}{M^*}\right) cr^d, \qquad \text{all} \quad x \in BG_1, \quad r \le r_-.$$

Thus, the supremum over $BG_1$ of the first term on the right-hand side of (9) is not greater than

$$\max\left(\frac{g_1 + \epsilon/M^*}{g_1 - 3\epsilon/M^*} - 1,\ 1 - \frac{g_1 - 2\epsilon/M^*}{g_1 + 3\epsilon/M^*}\right) \le \max\left(\frac{4\epsilon}{M^* g_1 - 3\epsilon},\ \frac{5\epsilon}{M^* g_1 + 3\epsilon}\right). \tag{10}$$

We can make (11) as small as desired by choosing $M^*$ large enough in view of the fact that $g_1 \ge \epsilon(1 - 1/M^*)$.

We conclude the proof of part 3 by applying Lemma 4 to the last term of (9). The number $b$ in Lemma 4 is equal to $\sup_x f(x)\, c(2r_-)^d$ (which is smaller than $1/4$ for all $n$ large enough). Thus,

$$P\left(\sup_{\substack{x \\ r \le r_-}} \frac{n}{k}\left|\mu_n(S_{xr}) - \mu(S_{xr})\right| > \epsilon\right)$$

$$\le 8(2n)^{d+1} \exp\left(-\frac{n(\epsilon k/n)^2}{64b + 4\epsilon k/n}\right) + 8n \exp\left(-\frac{nb}{10}\right) \tag{11}$$

for all $n$ so large that $nb \ge 1$ and $n(\epsilon k/n)^2 \ge 8b$. Now, $b = ak/n$ for some constant $a > 0$, so that the condition $\lim_{n\to\infty} k = \infty$ is sufficient for the applicability of the inequality of Lemma 4 for all $n$ large enough. The right-hand side of (11) does not exceed $c_1 n^{d+1} \exp(-c_2 k)$ for some positive constants $c_1$, $c_2$. When $k/\log n \to \infty$, this is in turn bounded by $\exp(-c_3 k)$ for all $n$ large enough, and some constant $c_3 > 0$. This concludes the proof of the theorem. Q.E.D.

*Proof of Theorem 2.* Assume first that Theorem 1 holds for all kernels $K$ of the form

$$K(x) = (ct^d)^{-1} I_{\|x\| \le t}, \qquad 0 < t \le 1. \tag{12}$$

Then we argue as follows: for fixed $\epsilon > 0$ find constants $N$, $a_1, \ldots, a_N$, and $u_1, \ldots, u_N$, all positive, such that the function $L^*$ defined by

$$L^*(u) = \sum_{j=1}^{N} a_j I_{0 < u \le u_j}$$

satisfies the inequality $L(u) \le L^*(u) \le L(u) + \epsilon$, all $0 \le u \le 1$. Note that $L^*(u) = 0$ outside $[0, 1]$. When we define

$$f_{nj}(x) = \frac{1}{n} \sum_{i=1}^{n} (cH_{ni}^d u_j^d)^{-1} I_{\|X_i - x\| \le u_j H_{ni}},$$

then

$$\begin{aligned} |f_n(x) - f(x)| &\le \left| \sum_{j=1}^{N} ca_j u_j^d f_{nj}(x) - f(x) \right| + \left| \sum_{j=1}^{N} ca_j u_j^d f_{nj}(x) - f_n(x) \right| \\ &\le \sum_{j=1}^{N} ca_j u_j^d |f_{nj}(x) - f(x)| + \left| \sum_{j=1}^{N} ca_j u_j^d - 1 \right| f(x) \\ &\quad + \frac{1}{n} \sum_{i=1}^{n} H_{ni}^{-d} \left| L\left(\left\| \frac{X_i - x}{H_{ni}} \right\|\right) - L^*\left(\left\| \frac{X_i - x}{H_{ni}} \right\|\right) \right|. \end{aligned} \tag{13}$$

The first term on the right-hand side of (13) satisfies (5) by assumption. The second term is not greater than $c \sup_x f(x)$ times

$$\left| \int_0^1 du^{d-1} L^*(u)\, du - \int_0^1 du^{d-1} L(u)\, du \right| \le \int_0^1 du^{d-1} |L^*(u) - L(u)|\, du \le \epsilon.$$

The last term does not exceed $c\epsilon f_n(x)$, where $f_n$ is as defined by (1) and (2). But by Theorem 1, $\sup_x \epsilon f_n(x) \le \epsilon \sup_x f(x) + \epsilon \sup_x |f_n(x) - f(x)|$, which is the sum of a term that can be made arbitrarily small by choice of $\epsilon$, and a term satisfying (5). Thus, Theorem 2 is proved if we can show it for kernels of the form (12).

The restriction $t \le 1$ allows us to continue using the convenient Lemma 2 in which $H_{ni}$ is replaced by $tH_{ni}$. In the expression for $U_{n1}$, $U_{n2}$, $U_{n3}$ in the proof of Theorem 1, we replace the indicator functions by

$$\frac{1}{t^d} I_{\|X_i - x\| \le tH_{ni}}.$$

Parts 1 and 2 remain unaffected except for a factor $1/t^d$ in the inequalities derived there. Part 3 too needs few changes; the most crucial one is the replacement of $k$ in (8) [the expression for $V_{ni}(x)$] by $kt^d$, with the same replacement throughout the remainder of part 3. Q.E.D.

## REFERENCES


Abou-Jaoude, S., (1974). Sur un théorème de non existence d'estimateurs convergeant en probabilité. *C. R. Acad. Sci. Paris, Sér. A*, 278, 1445–1448.

Breiman, L.; Meisel, W., and Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, 19, 135–144.

Chow, Y.S.; Geman, S., and Wu, L. (1982). Consistent cross-validated density estimation. Manuscript, Brown University, Providence, R.I.

Deheuvels, P. (1977a). Estimation non paramétrique de la densité par histogrammes géneralisés. *Rev. Statist. Appl.*, 25, 5–42.

Deheuvels, P. (1977b). Estimation non paramétrique de la densité par histogrammes géneralisés. *Publ. Inst. Statist. Univ. Paris*, 22, 1–23.

Devroye, L., and Wagner, T.J. (1977). The strong uniform consistency of nearest neighbor density estimates. *Ann. Statist.*, 5, 536–540.

Devroye, L., and Wagner, T.J. (1980). The strong uniform consistency of kernel density estimates. *Multivariate Analysis V* (P. R. Krishnaiah, *ed.*). North Holland, 59–77.

Devroye, L. (1985). A note on the $L_1$ consistency of variable kernel estimates. *Ann. Statist.*, 13, 1041–1049.

Devroye, L., and Gyorfi, L. (1985). *Nonparametric Density Estimation: The $L_1$ View*. Wiley, New York.

Duin, R.P.W. (1976). On the choice of smoothing parameters for Parzen Estimaters of probability density functions. *IEEE Trans. Comput.* C-25, 1175–1179.

Habbema, J.D.F.; Hermans, J., and Remme, J. (1978). Variable kernel density estimation in discriminant analysis. *COMPSTAT 1978* (L.C.A. Corsten and J. Hermans, *eds.*). Physica-Verlag, Wien, 178–185.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58, 13–30.

Loftsgaarden, D.O., and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist*, 36, 1049–1051.

Mack, Y.P., and Rosenblatt, M. (1979). Multivariate *k*-nearest neighbor density estimates. *J. Multivariate Anal.*, 9, 1–15.

Moore, D.S., and Yackel, J.W. (1977). Consistency properties of nearest neighbor density estimates. *Ann. Statist.*, 5, 143–154.

Nadaraya, E.A. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theory Probab. Appl.*, 19, 133–141.

Raatgever, J.W., and Duin, R.P.W. (1978). On the variable kernel model for multivariate nonparametric density estimation. *COMPSTAT 1978* (L.C.A. Corsten and J. Hermans, *eds.*). Physica-Verlag, Wien.

Rudemo, M. (1982). Empiricial choice of histograms and kernel density estimators. *Scand. J. Statist.*, 9, 65–78.

Schuster, E.F., and Gregory, G.G. (1981). On the nonconsistency of maximum likelihood nonparametric density estimators. *Computer Science and Statistics, 13th Symposium on the Interface* (W.F. Eddy, *ed.*). Springer-Verlag, New York, 295–298.

Scott, D.W., and Factor, L.E. (1981). Monte-Carlo study of three data-based nonparametric probability density estimators. *J. Amer. Statist. Assoc.*, 76, 9–15.

Scott, D.W.; Tapia, R.A., and Thompson, J.R. (1977). Kernel density estimation revisited. *Nonlinear Anal.*, 1, 339–372.

Silverman, B.W. (1978). Choosing the window when estimating a density. *Biometrika*, 65, 1–11.

Tapia, R.A., and Thompson, J.R. (1978). *Nonparametric Probability Density Estimation*. Johns Hopkins Univ. Press, Baltimore.

Wagner, T.J. (1975). Nonparametric estimates of probability densities. *IEEE Trans. Inform. Theory*, IT-21, 438–440.

Wertz, W. (1978). *Statistical Density Estimation: A Survey*. Applied Statistics and Econometrics Series, Volume 13. Vandenhoeck and Ruprecht, Göttingen.

Wertz, W., and Schneider, B. (1979). Statistical density estimation: A bibliography. *Internat. Statist. Rev.* 47, 155–175.

*Received 28 June 1985*
*Revised 27 February 1986*
*Accepted 3 March 1986*

*School of Computer Science*
*McGill University*
*Burnside Hall, 805 Sherbrooke Street West*
*Montréal, Quebec H3A 2K6*

*Applied Research Laboratories*
*The University of Texas at Austin*
*Austin, Texas 78712-8029*
*U.S.A.*