The Annals of Statistics
1977, Vol. 5, No. 3, 536-540

THE STRONG UNIFORM CONSISTENCY OF NEAREST NEIGHBOR DENSITY ESTIMATES

BY LUC P. DEVROYE AND T. J. WAGNER[1]

The University of Texas at Austin

Let $X_1, \cdots, X_n$ be independent, identically distributed random vectors with values in $\mathbb{R}^d$ and with a common probability density f. If $V_k(x)$ is the volume of the smallest sphere centered at x and containing at least k of the $X_1, \cdots, X_n$ then $f_n(x) = k/(nV_k(x))$ is a nearest neighbor density estimate of f. We show that if $k = k(n)$ satisfies $k(n)/n \to 0$ and $k(n)/\log n \to \infty$ then $\sup_x |f_n(x) - f(x)| \to 0$ w.p. 1 when f is uniformly continuous on $\mathbb{R}^d$.

Introduction. Suppose that $X_1, \cdots, X_n$ are independent, identically distributed random vectors with values in $\mathbb{R}^d$ and with a common probability density f. If $V_k(x)$ is the volume of the smallest sphere centered at x and containing at least k of the random vectors $X_1, \cdots, X_n$, then Loftsgaarden and Quesenberry (1965), to estimate $f(x)$ from $X_1, \cdots, X_n$, let

$$f_n(x) = k/(nV_k(x)) \tag{1}$$

where $k = k(n)$ is a sequence of positive integers satisfying

(2) (a) $$k(n) \uparrow \infty$$
(b) $$k(n)/n \to 0 .$$

(The factor $k - 1$ was used instead of k by Loftsgaarden and Quesenberry; this has no effect on any of the asymptotic results stated here.) They showed that $f_n(x)$ is a consistent estimate of $f(x)$ at each point where f is continuous and positive. This result can also easily be inferred from the work of Fix and Hodges (1951). For $d = 1$, Moore and Henrichon (1969) showed that

$$\sup_x |f_n(x) - f(x)| \to 0 \quad \text{in probability}$$

if f is uniformly continuous and positive on $\mathbb{R}$ and if, additionally,

$$k(n)/\log n \to \infty . \tag{3}$$

Wagner (1973) showed that $f_n(x)$ is a strongly consistent estimate of $f(x)$ at each continuity point of f if, in addition to (2b),

$$\sum_1^\infty e^{-\alpha k(n)} < \infty \qquad \text{for all} \quad \alpha > 0 . \tag{4}$$

(Notice that (4) is always implied by (3) but (2a) and (4) are needed to imply (3).) The result of this paper is the following theorem.

Received February 1976; revised October 1976.

[1] Research of the authors was sponsored by AFOSR GRANT 72-2371.

AMS 1970 *subject classifications.* 60F15, 62G05.

Key words and phrases. Nonparametric density estimation, multivariate density estimation, uniform consistency, consistency.

THEOREM. *If f is uniformly continuous on $\mathbb{R}^d$ and if $k(n)$ satisfies* (2b) *and* (3) *then*

$$\sup_x |f_n(x) - f(x)| \to_n 0 \quad \text{w.p. } 1 .$$

If

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

where K is the uniform probability density for the unit sphere in $\mathbb{R}^d$ and $\{r(n)\}$ is a sequence of positive numbers, the recent results of Moore and Yackel (1977) (see Theorem 3.1) and the above theorem immediately yield that

$$\sup_x |\hat{f}_n(x) - f(x)| \to 0 \quad \text{w.p. } 1$$

whenever f is uniformly continuous on $\mathbb{R}^d$ and $r(n) \to 0$, $nr(n)^d/\log n \to \infty$. This fact, an improvement over the previously published convergence results for the kernel estimate with a uniform kernel (e.g., see Theorem 2.1 of Moore and Yackel (1977)), also is a special case of Theorem 4.9 of Devroye (1976) who proves the same statement for all kernels K which are bounded probability densities with compact support and whose discontinuity points have a closure with Lebesgue measure 0.

PROOF. To simplify notation we assume below that multiplications are always carried out before division. Let $\varepsilon > 0$ and choose $\delta > 0$ such that

$$|f(y) - f(x)| < \varepsilon/2$$

whenever x and y are within a sphere of volume δ. Deferring measurability arguments for the moment,

$$P\{\sup_x |f_n(x) - f(x)| > \varepsilon\} \\ = P\{\bigcup_x [V_k(x) < k/n(f(x) + \varepsilon)]\} + P\{\bigcup_{x:f(x)>\varepsilon} [V_k(x) > k/n(f(x) - \varepsilon)]\} .$$

The event $\bigcup_x [V_k(x) < k/n(f(x) + \varepsilon)]$ implies that, for some x, there must be a sphere centered at x with volume less than $k/n(f(x) + \varepsilon)$ and containing k of the random vectors $X_1, \cdots, X_n$. If $k/n\varepsilon < \delta$ then the probability measure of such a sphere must be less than $k(f(x) + \varepsilon/2)/n(f(x) + \varepsilon)$ so that, for one of these spheres S,

$$\mu_n(S) - \mu(S) > \frac{k}{n} - \frac{k(f(x) + \varepsilon/2)}{n(f(x) + \varepsilon)} \\ = \frac{k\varepsilon}{2n(f(x) + \varepsilon)} \geqq \frac{k\varepsilon}{2n(F + \varepsilon)}$$

where F is the maximum of f on $\mathbb{R}^d$, μ is the measure on the Borel subsets of $\mathbb{R}^d$ corresponding to f and μ_n is the empirical measure on the Borel subsets of $\mathbb{R}^d$ for $X_1, \cdots, X_n$. Thus, for $k/n\varepsilon < \delta$,

$$(5) \qquad P\{\bigcup_x [V_k(x) < k/n(f(x) + \varepsilon)]\} \\ \leqq P\{\sup_{S \in \mathscr{A}_n} |\mu_n(S) - \mu(S)| > k\varepsilon/2n(F + \varepsilon)\}$$

where $\mathscr{A}_n$ is the class of all spheres in $\mathbb{R}^d$ whose volume is less than $4k/n\varepsilon$.

Next, with $4k/n\varepsilon < \delta$,

$$\bigcup_{x:f(x)>\varepsilon}[V_k(x) > k/n(f(x)-\varepsilon)] \subseteq \bigcup_{x:f(x)>\varepsilon}[V_k(x) > k/n(f(x)-(3\varepsilon/4))]$$

which implies that, for some x with $f(x) > \varepsilon$, there is a sphere S centered at x, with volume $\leqq 4k/n\varepsilon$, and

$$\mu(S) \geqq k(f(x)-\varepsilon/2)/n(f(x)-(\tfrac{3}{4})\varepsilon)\,,$$
$$\mu_n(S) \leqq k/n\,, \qquad \text{and}$$
$$\mu(S)-\mu_n(S) \geqq k\varepsilon/4n(f(x)-(\tfrac{3}{4})\varepsilon)\,.$$

Thus

$$P\{\bigcup_{x:f(x)>\varepsilon}[V_k(x) > k/n(f(x)-\varepsilon)]\} \leqq P\{\sup_{S\in\mathscr{A}_n}|\mu(S)-\mu_n(S)| \geqq k\varepsilon/4nF\}\,, \tag{6}$$

so that

$$P\{\sup_x|f_n(x)-f(x)| \geqq \varepsilon\} \leqq 2P\{\sup_{S\in\mathscr{A}_n}|\mu_n(S)-\mu(S)| \geqq k\varepsilon/4n(F+\varepsilon)\}\,.$$

The proof will be completed if we show that for each $\varepsilon > 0$

$$\textstyle\sum_n P\{\sup_{S\in\mathscr{A}_n}|\mu_n(S)-\mu(S)| \geqq k\varepsilon/4n(F+\varepsilon)\} < \infty\,. \tag{7}$$

To prove (7) we employ a variation of the argument used by Vapnik and Chervonenkis (1971). In this variation use will be made of the following result. If $Y_1, \cdots, Y_n$ represent independent drawings without replacement from a population of k 0's and 1's then, for $\varepsilon > 0$ and $k \geqq n$,

$$P[|(\textstyle\sum_1^n Y_i)/n-\mu| \geqq \varepsilon] \leqq 2e^{-n\varepsilon^2/(2\mu+\varepsilon)} \tag{8}$$

where μ, the {number of 1's}/k, is assumed to be $\leqq \frac{1}{2}$. Additionally (8) holds when $Y_1, \cdots, Y_n$ are Bernoulli random variables with parameter $\mu \leqq \frac{1}{2}$. (Use the two-sided version of Theorem 3 of Hoeffding (1963) along with $\mu \leqq \frac{1}{2}$ and $\log(1+(\varepsilon/\mu)) \geqq 2\varepsilon/(2\mu+\varepsilon)$. See also Section 6 of this paper.)

Now, if $\sup_{\mathscr{A}}\mu(A) \leqq M$ and $n \geqq 8M/\delta^2$, an easy modification of Lemma 1 of Vapnik and Chervonenkis (1971) yields

$$P[\sup_{\mathscr{A}}|\mu_n(A)-\mu(A)| \geqq \delta] \leqq 2P[\sup_{\mathscr{A}}|\mu_n(A)-\mu_n'(A)| \geqq \delta/2] \tag{9}$$

where $\mu_n'(A)$ is the empirical measure for A with $X_{n+1}, \cdots, X_{2n}$ and $\mathscr{A}$ is any class of Borel sets in $\mathbb{R}^d$ for which

$$\sup_{\mathscr{A}}|\mu_n(A)-\mu(A)| \qquad \text{and} \qquad \sup_{\mathscr{A}}|\mu_n(A)-\mu_n'(A)|$$

are random variables. Putting $\mathscr{A} = \mathscr{A}_n$ we see that M can be taken to be $4kF/n\varepsilon$. Since, for $\alpha > 0$,

$$\begin{aligned} &P[\sup_{\mathscr{A}_n}|\mu_n(A)-\mu_n'(A)| \geqq \delta/2] \\ &\quad \leqq P[\sup_{\mathscr{A}_n}|\mu_n(A)-\mu_n'(A)| \geqq \delta/2;\ \sup_{\mathscr{A}_n}\mu_{2n}(A) \leqq \alpha M] \\ &\qquad + P[\sup_{\mathscr{A}_n}\mu_{2n}(A) > \alpha M] \end{aligned} \tag{10}$$

we see, using (3) and putting $\delta = k\varepsilon/4n(F+\varepsilon)$, that (7) follows whenever both terms of the right-hand side of (10) are summable for some $\alpha > 0$. Looking at the first term, we note that it equals

$$\int_{\mathbb{R}^{2nd}} \frac{1}{(2n)!} \sum I_{[\sup_{\mathscr{A}_n} |\mu_n(A)-\mu_n'(A)| \geqq \delta/2]} I_{[\sup_{\mathscr{A}_n} \mu_{2n}(A) \leqq \alpha M]} \, dQ$$

where I_E is the indicator of the set $E \subseteqq \mathbb{R}^d$ and Q is the probability measure on $\mathbb{R}^{2nd}$ for $X_1, \cdots, X_{2n}$ and where the inner summation is taken over all $(2n)!$ permutations of $x_1, \cdots, x_{2n}$. But this last integral equals

$$\int_{\mathbb{R}^{2nd}} \frac{1}{(2n)!} \sum I_{[\sup_{\mathscr{A}_n} \mu_{2n}(A) \leqq \alpha M]} \sup_{\mathscr{A}_n} I_{[|\mu_n(A)-\mu_n'(A)| \geqq \delta/2]} \, dQ$$

$$= \int_{\mathbb{R}^{2nd}} \frac{1}{(2n)!} \sum I_{[\sup_{\mathscr{A}_n} \mu_{2n}(A) \leqq \alpha M]} \sup_{\mathscr{A}'} I_{[|\mu_n(A)-\mu_n'(A)| \geqq \delta/2]} \, dQ$$

$$\leqq \int_{\mathbb{R}^{2nd}} \sum_{A \in \mathscr{A}'} I_{[\sup_{\mathscr{A}_n} \mu_{2n}(A) \leqq \alpha M]} \left\{ \frac{1}{(2n)!} \sum I_{[|\mu_n(A)-\mu_{2n}(A)| \geqq \delta/4]} \right\} dQ$$

where $\mathscr{A}' = \mathscr{A}'(x_1, \cdots, x_{2n})$ is any finite subclass of $\mathscr{A}_n$ which yields the same class of intersections with $\{x_1, \cdots, x_{2n}\}$ and where the inner summation is again taken over the $(2n)!$ permutations of $x_1, \cdots, x_{2n}$. The quantity within $\{\cdot\}$ is bounded above, using (8), by

$$2e^{-n\delta^2/(32\mu_{2n}(A)+4\delta)}$$

whenever $\mu_{2n}(A) \leqq \frac{1}{2}$. Since $M = 4kF/n\varepsilon$ we see, from (3), that for all n sufficiently large the last integral is upper-bounded by

$$2 \int_{\mathbb{R}^{2nd}} e^{-n\delta^2/(32\alpha M+4\delta)} \left(\sum_{A \in \mathscr{A}'} 1\right) dQ \, .$$

Choosing $\mathscr{A}'$ to be a smallest possible subclass, we have (Vapnik and Chervonenkis (1971), Cover (1965)) that $(\sum_{A \in \mathscr{A}'} 1) \leqq 1 + (2n)^{d+3}$ and, using (3) again, that the first term of (10) is summable for all $\alpha > 0$.

Looking at the second term of (10), let r be the radius of a sphere in $\mathbb{R}^d$ whose volume is $4k/n\varepsilon$. If some sphere of radius r contains l of the points $X_1, \cdots, X_{2n}$ then there must be at least one sphere or radius $2r$, centered at one of the points $X_1, \cdots, X_{2n}$ which contains at least l points. Thus

$$P[\sup_{\mathscr{A}_n} \mu_{2n}(A) > \alpha M] \leqq 2nP[\mu_{2n}(S_{X_1}(2r)) > \alpha M]$$

where $S_x(t)$ denotes the sphere of radius t centered at x. But

$$P[\mu_{2n}(S_{X_1}(2r)) > \alpha M]$$
$$\leqq \max_{x \in \mathbb{R}^d} P[\mu_{2n-1}(S_x(2r)) > (\alpha 2nM - 1)/(2n-1)]$$
$$\leqq \max_{x \in \mathbb{R}^d} P[\mu_{2n-1}(S_x(2r)) - \mu(S_x(2r)) > [(\alpha 2nM - 1)/(2n-1)] - 2^d 4kF/n\varepsilon] \, .$$

At this point it is not difficult, using (3) and (8), to show that the second term of (9) is summable as long as $\alpha > 2^d$.

Finally, to complete the proof, it is easy to see that all of the uncountable unions over x are indeed events and that the various supremums over $\mathscr{A}_n$ are indeed random variables.

REFERENCES

COVER, T. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Computers* **10** 326–334.

DEVROYE, L. P. (1976). Nonparametric discrimination and density estimation. Ph. D. thesis, Univ. of Texas, Austin.

FIX, E. and HODGES, J. L. (1951). Discriminatory analysis. Nonparametric discrimination: consistency properties. Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

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.

MOORE, D. S. and HENRICHON, E. G. (1969). Uniform consistency of some estimates of a density function. *Ann. Math. Statist.* **40** 1499–1502.

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

VAPNIK, V. N. and CHERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probability Appl.* **16** 264–280.

WAGNER, T. J. (1973). Strong consistency of a nonparametric estimate of a density function. *IEEE Trans. Systems, Man, and Cybernet.* **3**, 289–290.

DEPARTMENT OF ELECTRICAL ENGINEERING
THE UNIVERSITY OF TEXAS AT AUSTIN
COLLEGE OF ENGINEERING
AUSTIN, TEXAS 78712