The Annals of Statistics
1979, Vol. 7, No. 5, 1136–1139

THE L_1 CONVERGENCE OF KERNEL DENSITY ESTIMATES[1]

BY L. P. DEVROYE AND T. J. WAGNER

University of Texas

Let $X_1, \cdots, X_n$ be a sequence of independent random vectors taking values in $\mathbb{R}^d$ with a common probability density f. If $f_n(x) = (1/h) h_n^{-d}\Sigma_{i=1}^n K((x - X_i)/h_n)$ is the kernel estimate of f from $X_1, \cdots, X_n$ then conditions on K and $\{h_n\}$ are given which insure that $\int |f_n(x) - f(x)|\, dx \to_n 0$ in probability or with probability one. No continuity conditions are imposed on f.

Let $X_1, \cdots, X_n$ be a sequence of independent random vectors taking values in $\mathbb{R}^d$ with a common probability density f. The kernel estimate of f from $X_1, \cdots, X_n$ is given by

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

where the kernel K is a bounded probability density on $\mathbb{R}^d$ and $\{h_n\}$ is a sequence of positive numbers. We are concerned here with the conditions of f, K and $\{h_n\}$ which insure the L_1 convergence of f_n to f, namely,

$$(1) \qquad \int_{\mathbb{R}^d}|f_n(x) - f(x)|\, dx \to_n 0 \qquad \text{in probability (or w.p. 1).}$$

This concern is motivated by the observation (Scheffé (1947)) that

$$2 \sup_{B \in \mathcal{B}}|\mu_n(B) - \mu(B)| = \int_{\mathbb{R}^d}|f_n(x) - f(x)|\, dx$$

where $\mathcal{B}$ is the class of Borel sets in $\mathbb{R}^d$ and μ_n and μ are the measures on $\mathcal{B}$ corresponding to f_n and f respectively. Consequently, whenever (1) holds

$$(2) \qquad \sup_{B \in \mathcal{B}}|\mu_n(B) - \mu(B)| \to_n 0 \qquad \text{in probability (or w.p. 1).}$$

Of course, (2) reminds one of the Glivenko-Cantelli theorem and its extensions (Winter (1973), Glick (1974)), namely,

$$\sup_{B \in \mathcal{C}}|\nu_n(B) - \nu(B)| \to_n 0 \qquad \text{in probability (or w.p. 1),}$$

where $X_1, \cdots, X_n$ are independent, identically distributed with an arbitrary probability measure ν on $\mathcal{B}$, ν_n is the empirical measure for $X_1, \cdots, X_n$ and $\mathcal{C}$ is a strict subclass of Borel sets (Rao (1962), Vapnik and Chervonenkis (1971)). For our case it is easy to see that μ must be absolutely continuous if (2) is to hold for μ_n which correspond to kernel estimates.

Glick (1974) has shown that whenever f_n is a probability density on $\mathbb{R}^d$ which is a measurable function of x and $X_1, \cdots, X_n$, then (1) follows from

$$(3) \qquad f_n(x) \to_n f(x) \qquad \text{in probability (or w.p. 1) almost everywhere in } x.$$

Received May 1977; revised January 1978.
[1]Research supported by AFOSR Grant 77-3385.
AMS 1970 *subject classifications.* 60F15, 62G05.
Key words and phrases. Density estimation, integral convergence, kernel estimates.

Whenever f is almost everywhere continuous on $\mathbb{R}^d$, (3) then follows immediately from the known pointwise consistency conditions for kernel estimates. (See, for example, Rosenblatt (1957), Parzen (1962), Cacoullos (1965), Nadaraya (1965), Van Ryzin (1969), Deheuvels (1974).) This argument fails for those densities on $\mathbb{R}^d$ which do not have an almost everywhere continuous version. The main result of this note is that (1) follows without any continuity requirements on f and, consequently, (2) holds for all absolutely continuous probability measures.

For comparison, we note that the nearest neighbor density estimate of f (Loftsgaarden and Quesenberry (1965), Moore and Yackel (1977)) will never satisfy (1) or (2) since its integral over $\mathbb{R}^d$ is always infinite. Abou-Jaoude (1976a, 1976b) has shown, however, that (1) holds for different types of histogram estimates with no assumptions on f.

THEOREM. *Let K be a bounded probability density on $\mathbb{R}^d$ with*

$$L(u) = \sup_{\|x\| \geqslant u} K(x)$$

for $u \geqslant 0$. If $\{h_n\}$ is a sequence of positive numbers then (1) *follows whenever*

$$h_n \to_n 0 \tag{4}$$

$$nh_n^{\,d} \to_n \infty \ \left(\Sigma_1^\infty e^{-\alpha n h_n^{\,d}} < \infty \text{ for all } \alpha > 0\right) \tag{5}$$

and one of the following conditions holds:

$$\|x\|^d K(x) \to 0 \text{ as } \|x\| \to \infty \text{ and } f \text{ is almost everywhere continuous,} \tag{6}$$

$$f \text{ is bounded,} \tag{7}$$

$$\textstyle\int_0^\infty u^{d-1} L(u)\, du < \infty. \tag{8}$$

REMARK. The condition in (6) imposed on K is equivalent to

$$u^d L(u) \to 0 \qquad \text{as} \quad u \to \infty$$

which is only slightly weaker than (8).

PROOF. Starting, as usual, with

$$|f_n(x) - f(x)| \leqslant |f_n(x) - Ef_n(x)| + |Ef_n(x) - f(x)|$$

we first show that

$$Ef_n(x) \to_n f(x) \qquad \text{almost everywhere in } x. \tag{9}$$

The usual argument shows that (9) is implied by (4) and (6) (e.g., use the d-dimensional version of Theorem 1A of Parzen (1962)). Next

$$|Ef_n(x) - f(x)| \leqslant \textstyle\int_{\|y\| < \delta h_n} |f(x-y) - f(x)| h_n^{-d} K(y/h_n)\, dy \\ + \textstyle\int_{\|y\| \geqslant \delta h_n} |f(x-y) - f(x)| h_n^{-d} K(y/h_n)\, dy. \tag{10}$$

If $\lambda(B)$ denotes the Lebesgue measure of the Borel set $B \subseteq \mathbb{R}^d$ and if $S(x, r)$ denotes the closed sphere of radius r centered at x then the first term of the right-hand side of (10) is bounded by

$$\sup_y K(y) \lambda(S(0, \delta)) \textstyle\int_{S(x, \delta h_n)} \{|f(y) - f(x)| / \lambda(S(x, \delta h_n))\}\, dy$$

which tends to 0 for almost every x and every $\delta > 0$ if $h_n \to_n 0$ (see, for example, Zygmund (1959, 1969)). If f is bounded the second term of (10) is bounded by

$$2 \sup_y f(y) \int_{\|y\|>\delta} K(y)\, dy$$

which can be made arbitrarily small for all n by taking δ large enough. Thus (4) and (7) imply (9). Using a theorem of Stein ((1970), pages 62–63) we see that (4) and (8) imply (9).

Looking at $f_n(x) - Ef_n(x)$ we see that it equals

$$\frac{1}{n}\Sigma_1^n(Y_{ni} - EY_{ni})$$

where

$$Y_{ni} = h_n^{-d}K((x - X_i)/h_n).$$

Letting $\sup_y K(y) = M$, we have

$$0 \leqslant Y_{ni} \leqslant M/h_n^{\ d},$$

and

$$EY_{ni}^{\ 2} \leqslant (M/h_n^{\ d})Ef_n(x),$$

so that, by Bennett's inequality (Bennett (1962)),

$$P\{|f_n(x) - Ef_n(x)| \geqslant \varepsilon\} \leqslant \exp(-2n\varepsilon^2 h_n^{\ d}/(2MEf_n(x) + M\varepsilon)).$$

At each point x for which $Ef_n(x) \to_n f(x)$ the sequence $\{Ef_n(x)\}$ remains bounded so that, almost everywhere in x,

$$f_n(x) - Ef_n(x) \to_n 0 \qquad \text{in probability or w.p. 1}$$

depending on whether $nh_n^d \to_n \infty$ or $\Sigma_1^\infty e^{-\alpha n h_n^d} < \infty$ for all $\alpha > 0$. Since (3) follows from the conditions of the theorem, (1) now follows from Glick's result.

REMARK. The proof also yields the strong pointwise consistency of f_n whenever K is a bounded probability density and

(i) $h_n \to_n 0$,
(ii) $\Sigma_1^\infty e^{-\alpha n h_n^d} < \infty$ for $\alpha > 0$, and
(iii) $\|x\|^d K(x) \to 0$ as $\|x\| \to \infty$ or f is bounded.

This result is similar to the one obtained by Deheuvels (1974).

Acknowledgment. We wish to thank the referee for pointing out a nice improvement in an earlier version of the theorem given here.

REFERENCES

ABOU-JAOUDE, S. (1976a). Conditions nécessaires et suffisantes de convergence L_1 en probabilité de l'histogramme pour une densité. *Ann. Inst. H. Poincaré Sect. B* **12** 213–231.

ABOU-JAOUDE, S. (1976b). Sur la convergence L_1 et L of l'estimateur de la partition aléatoire pour une densité. *Ann. Inst. H. Poincaré Sect. B* **12** 299–317.

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

CACOULLOS, T. (1965). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 179–190.

DEHEUVELS, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C. R. Acad. Sci. Paris Ser. A.* **278** 1217–1220.

GLICK, N. (1974). Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Math.* **6** 61–74.

LOFTSGAARDEN, D. O. and QUESENBERRY, C. P. (1965). A nonparametric estimate of a probability density. *Ann. Math. Statist.* **36** 1049–1051.

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

NADARAYA, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theor. Probability Appl.* **10** 186–190.

PARZEN, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* **33** 1065–1076.

RAO, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

ROSENBLATT, M. (1957). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–458.

STEIN, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press, Princeton, N.J.

VAN RYZIN, J. (1969). On the strong consistency of density estimates. *Ann Math. Statist.* **40** 1765–1772.

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

WINTER, B. B. (1973). Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.* **1** 247–253.

ZYGMUND, A. (1959). *Trigonometric Series, Vols. I and II.* Cambridge Univ. Press.

ZYGMUND, A. (1969). On certain lemmas of Marcinkiewicz and Carleson. *J. Approximation Theory* **2** 249–257.

DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITY OF TEXAS
P.O. BOX 7728
AUSTIN, TEXAS 78712