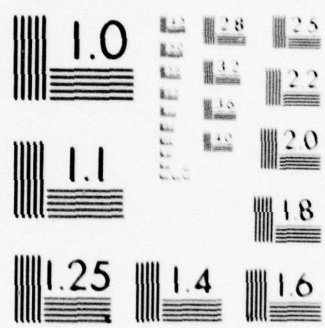
AD-A072 133   FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS   F/G 12/1
CONVERGENCE RATES FOR THE MEAN INTEGRATED SQUARED ERRORS OF SOM--ETC(U)
JUL 79   E P CHENG, R J SERFLING   N00014-76-C-0608

UNCLASSIFIED   FSU-STATISTICS-M505   NL

1 OF 1

AD
A072133

END
DATE
FILMED
9-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

(6)

CONVERGENCE RATES FOR THE MEAN INTEGRATED SQUARED ERRORS OF SOME NONPARAMETRIC DENSITY ESTIMATORS OF RECURSIVE δ-FUNCTION TYPE

Delta

~~by~~ K. F. Cheng ~~and~~ R. J. Serfling

(10)

~~FSU Statistics Report M505~~
~~ONR~~ Technical Report, ~~No. 138~~

(9)

(12) 15p.

(11) July, 1979

Department of Statistics
The Florida State University
Tallahassee, Florida 32306

(14)

FSU-STATISTICS-M5Ø5, TR-138-ONR

(15)

Research supported by ~~the Army, Navy and~~ Air Force under Office of Naval Research Contract No. N00014-76-C-0608. Reproduction in whole or in part is permitted for any ~~purpose of the United~~ States Government.

400 277

ABSTRACT

CONVERGENCE RATES FOR THE MEAN INTEGRATED SQUARED ERRORS OF SOME NONPARAMETRIC DENSITY ESTIMATORS OF RECURSIVE δ-FUNCTION TYPE

For estimation of a probability density function f by an empirical function f_n based on a sample of size n from f, a widely used measure of goodness is the mean integrated squared error. For the well known δ-function type of f_n, we show that the asymptotic behavior of this measure is essentially unchanged if f_n is replaced by a recursive version. Also, we characterize this asymptotic behavior under somewhat milder smoothness restrictions on f than previously considered in the literature, at the expense however of adding tail restrictions on f.

Key phrases: Nonparametric density estimation; mean integrated squared error; convergence rates; recursive estimators.

1. Introduction and a basic lemma. Consider estimation of an unknown probability density function f by an empirical function f_n based on a sample. A widely used measure of goodness is the mean integrated squared error. For the well-known δ-function type of f_n, we show that the asymptotic behavior of this measure is essentially unchanged if f_n is replaced by a recursive version $\tilde{f}_n$. Also, we characterize the asymptotic behavior under somewhat milder conditions on f than previously considered in the literature, and compare with previous results.

Specifically, we consider "nonparametric" estimation of a density f ($f(x) \geq 0$, $\int f(x)dx = 1$) on the real line by an empirical function f_n (not necessarily itself a density) formed from a sample of independent random variables $X_1, \ldots, X_n$ each having density f. Various ways of concocting a useful f_n are reviewed in Wegman (1972), Fryer (1977), and Tapia and Thompson (1978). Here we consider the *δ-function* (or in specialized form the *kernel* type) approach of Rosenblatt (1956), Whittle (1958), Parzen (1962), Leadbetter (1963), Watson and Leadbetter (1963), and Nadaraya (1974), in which f_n has the form

$$f_n(x) = n^{-1} \sum_{i=1}^{n} \delta_n(x - X_i), \quad -\infty < x < \infty, \tag{1.1}$$

where $\delta_n(\cdot)$ is a suitably chosen approximant to the Dirac δ-function. The kernel type f_n corresponds to δ_n of the form

$$\delta_n(u) = c_n^{-1} K(u/c_n), \quad -\infty < u < \infty,$$

where $K(\cdot)$ is a specified "kernel" or weight function and $\{c_n\}$ is a sequence of specified "bandwidth" constants tending to 0.

A useful global measure of goodness of an estimator f_n for f is the *mean integrated squared error* (MISE),

$$J(f, f_n) = E\{\int [f_n(x) - f(x)]^2 dx\}.$$

Under restrictions on f, and for f_n chosen compatibly with respect to such restrictions, the rate of convergence of $J(f, f_n)$ to 0 has been characterized by various authors. For δ-function f_n, Watson and Leadbetter (1963) impose conditions on the characteristic function of f. For kernel type f_n, Nadaraya (1974) imposes conditions on the derivatives of f of order 2 and higher. The validity of such conditions may presuppose more knowledge of the unknown f than would be available realistically, thus making difficult the selection of a "compatible" f_n. In Section 2 we offer some competitive results under conditions on the tails of f (implied by simple moment conditions). Also, we extend Nadaraya's approach to the case of restrictions merely on the first derivative of f.

A variation on (1.1) is the associated *recursive* version defined by

$$\tilde{f}_n(x) = n^{-1} \sum_{i=1}^{n} \delta_i(x - X_i), \tag{1.2}$$

introduced in the context of kernel type f_n by WoLverton and Wagner (1969). We have $\tilde{f}_n(x) = n^{-1}[(n - 1)\tilde{f}_{n-1}(x) + \delta_n(x - X_n)]$, so that the estimator need not be recomputed entirely when an additional observation is combined with previous ones. Besides saving computational effort, this lends itself to sequential sampling (see Davies and Wegman (1975)). On the other hand, there is the philosophical grievance that in (1.2), versus (1.1), the observations X_i are not being used symmetrically.

Nevertheless, for large samples, which is the case when the recursive approach is of greatest importance, this asymmetry does not affect efficiency, as we will see below. Also, ((1.2) is structurally simpler than (1.1), in that (1.2) involves partial sums over the *single sequence* of random variables $\{\delta_n(x - X_n), n \geq 1\}$, whereas (1.1) entails the *double array* $\{\delta_n(x - X_i), 1 \leq i \leq n, n \geq 1\}$. Thus the application of classical probability theory is more straightforward and fruitful in connection with (1.2). For example, it is easy to state, pointwise in x, a law of the iterated logarithm for the estimator $\tilde{f}_n(x)$.

For kernel type f_n, the work of Parzen (1962) on pointwise mean square error of $f_n(x)$ has been extended to recursive versions by Yamato (1971) and Wegman and Davies (1979). Winter (1978) has investigated a global measure, the *uniform mean square error* $\sup_x E\{[\tilde{f}_n(x) - f(x)]^2\}$. However, the MISE has not received attention. We now prove, for general δ-function estimators, a simple lemma which yields the relation

$$(1.3) \qquad \overline{\lim}_{n\to\infty} J(f, \tilde{f}_n) \leq \overline{\lim}_{n\to\infty} J(f, f_n),$$

showing that the recursive version $\tilde{f}_n$ is asymptotically as effective as the nonrecursive counterpart. In proving the lemma, we will use the fact that by Fubini's theorem the MISE may be written as the "IMSE," which in turn may be written as the sum of the *integrated variance* and the *integrated squared bias*. That is, for any estimator f_n, we have

$$(1.4) \qquad J(f, f_n) = A(f, f_n) + B(f, f_n),$$

where $A(f, f_n) = \int E\{[f_n(x) - Ef_n(x)]^2\}\, dx$ and $B(f, f_n) = \int [Ef_n(x) - f(x)]^2 dx$.

LEMMA. *For a given δ-function sequence* $\{\delta_n\}$, *let* f_n *and* $\tilde{f}_n$ *be given by* (1.1) *and* (1.2), *respectively. Then*

$$J(f, \tilde{f}_n) \le n^{-1} \sum_{i=1}^{n} J(f, f_i). \tag{1.5}$$

PROOF. We have

$$\begin{aligned} E\{[\tilde{f}_n(x) - E\tilde{f}_n(x)]^2\} &= \mathrm{Var}\{n^{-1} \sum_{n=1}^{n} \delta_i(x - X_i)\} \\ &= n^{-2} \sum_{i=1}^{n} \mathrm{Var}\{\delta_i(x - X_i)\} \\ &= n^{-2} \sum_{i=1}^{n} i\ \mathrm{Var}\{f_i(x)\} \\ &\le n^{-1} \sum_{i=1}^{n} \mathrm{Var}\{f_i(x)\}. \qquad (1.6) \end{aligned}$$

And, by Jensen's inequality,

$$\begin{aligned} [E\tilde{f}_n(x) - f(x)]^2 &= \{n^{-1} \sum_{i=1}^{n} [E\delta_i(x - X_i) - f(x)]\}^2 \\ &\le n^{-1} \sum_{i=1}^{n} [E\delta_i(x - X_i) - f(x)]^2 \\ &= n^{-1} \sum_{i=1}^{n} [Ef_i(x) - f(x)]^2 \qquad (1.7) \end{aligned}$$

By (1.6) and (1.7) we have $A(f, \tilde{f}_n) \le n^{-1}\sum_{i=1}^{n} A(f, f_i)$ and $B(f, \tilde{f}_n) \le n^{-1}\sum_{i=1}^{n} B(f, f_i)$, from which (1.5) follows by 1.4). □

Thus $J(f, \tilde{f}_n)$ is dominated by the Cesàro mean of $\{J(f, f_n)\}$ and hence (1.3) follows.

2. <u>On rates of convergence for the MISE</u>. To establish a perspective, let us first review three key results in the literature. Watson and Leadbetter (1963) consider two forms of restriction on the rate of decrease

of the characteristic function ϕ_f of the underlying density f. Either ϕ_f has an *exponential* rate with coefficient $\rho > 0$, i.e.,

(2.1a) $$|\phi_f(t)| \le Ae^{-\rho|t|}, \text{ all } t, \text{ for some constant } A,$$

or it has an *algebraic* rate of degree $p > 0$, i.e.,

(2.2a) $$\lim_{|t|\to\infty} |t|^p |\phi_f(t)| < \infty.$$

In this connection two types of δ-sequence $\{\delta_n\}$ for estimators given by (1.1) are considered. In each case δ_n is assumed square integrable with Fourier transform ϕ_{δ_n}. Either $\phi_{\delta_n}(t)$ has the form

(2.1b) $$h(a_n e^{\alpha|t|})$$

or it has the form

(2.2b) $$h(a_n t),$$

where in either case h is a bounded square integrable function and $\{a_n\}$ is a sequence of constants tending to 0. In the first case the auxiliary conditions

(2.1c) $$|1 - h(t)| \le B|t|, \ |t| \le 1, \text{ for some constant } B,$$

and

(2.1d) $$a_n = Dn^{-b}, \text{ for } b > \tfrac{1}{2} \text{ and } b \ge \alpha/2\rho,$$

are required, and in the second case the auxiliary conditions

(2.2c) $$\int |t|^{-2p}[1 - h(t)]^2 dt < \infty$$

and

(2.2d) $$a_n = Dn^{-1/2p}.$$

In this setting Watson and Leadbetter (1963) establish: *under conditions* (2.1),

(A) $$\lim_{n\to\infty}(n/\log n)J(f, f_n) = Q_A;$$

under conditions (2.2),

(B) $$\lim_{n\to\infty} n^{1-1/2p} J(f, f_n) = Q_B.$$

Here Q_A and Q_B are specified finite constants whose values are not important in the present discussion.

We mention that (2.2b) equivalently means that $\delta_n(u) = a_n^{-1}K(u/c_n)$, with $\phi_K(t) = h(t)$, making f_n of kernel type with square integrable kernel K. In any case, in order to assert for an estimator f_n the rate (A) or (B) for $J(f, f_n)$, one must construct f_n so as to meet conditions (2.1b, c, d) or (2.2b, c, d). This requries prior knowledge of (2.1a) or (2.2a), respectively, for the unknown density f.

Alternatively, Nadaraya (1974) utilizes conditions on the derivatives of f. For even integer $s \geq 2$, let W_s denote the set of functions f(x) having derivatives of s-th order with $f^{(s)}(x)$ being a bounded continuous $L_2(-\infty, \infty)$ function, and let H_s denote the class of kernels K(u) satisfying $K(u) = K(-u)$, $\int K(u)du = 1$, $\sup_u |K(u)| < \infty$, $\int u^i K(u)du = 0$, $1 \leq i \leq s - 1$, $\int u^s K(u)du \neq 0$, and $\int u^s |K(u)|du < \infty$. For estimation of f satisfying

(2.3a) $$f \in W_s,$$

the "compatible" kernel type estimator f_n corresponds to $\delta_n(u) = c_n^{-1}K(u/c_n)$ with

(2.3b) $$K \in H_s$$

and

$$c_n = Cn^{-1/(2s+1)}. \tag{2.3c}$$

Nadaraya (1974) establishes: *under conditions* (2.3),

$$\lim_{n\to\infty} n^{1-1/(2s+1)} J(f, f_n) = Q_C. \tag{C}$$

This result is competitive with (B), but for $s \geq 2$ the condition (2.3a) presupposes considerable knowledge of the properties of f. We now establish two results requiring less stringent prior information. We restrict to kernel type f_n.

THEOREM 1. *Let f have continuous squre integrable derivative. Let K satisfy* $\int K(u)du = 1$ *and* $\int u^2|K(u)|du < \infty$, *and take* $c_n = Cn^{-1/3}$. *Then*

$$J(f, f_n) = O(n^{-2/3}). \tag{D}$$

PROOF. Using (1.4), we treat the terms $A(f, f_n)$ and $B(f, f_n)$ separately. Assuming K square integrable and requiring nothing of f, we have

$$A(f, f_n) = n^{-1}\int \mathrm{Var}\{\delta_n(x - X_1)\}dx \leq n^{-1}\int E\{\delta_n^2(x - X_1)\}dx$$

$$= (nc_n)^{-1}\int K^2(u)du. \tag{2.4}$$

In dealing with the term $B(f, f_n)$, we apply Taylor's formula with integral form of remainder, to write

$$Ef_n(x) - f(x) = \int K(u)[f(x - c_n u) - f(x)]du$$

$$= c_n\int K(u) \int_0^1 f'(x - tc_n u)u \, dt \, du,$$

giving

$$B(f, f_n) \le c_n^2\int[\int|K(u)|\int_0^1|f'(x - tc_n u)u\,dt\,du]^2dx$$

$$\le c_n^2\int\int[(\int_0^1|f'(x - tc_n u)dt)^2u^2|K(u)|du]dx$$

$$= c_n^2(\int[f'(x)]^2dx)(\int u^2|K(u)|du), \tag{2.5}$$

by two applications of Jensen's inequality. Combining (2.4) and (2.5) with $c_n = Cn^{-1/3}$, (D) follows. □

The preceding result extends the range of Nadaraya's conditions on f down to first order derivative restrictions. We have, in fact, used Nadaraya's line of argument in the above proof. In the next result we introduce a modification in the treatment of $B(f, f_n)$ and further relax the conditions on smoothness of f, but at the expense of adding restrictions on the tails of f. For each real $q > 0$, we put

$$C_{f,q} = \sup_t t^q \int_{|x|>t} f(x)dx \quad (\le \int|x|^q f(x)dx).$$

Also, put

$$C_{f,\infty} = \sup\{t: \int_{|x|\le t} f(x)dx = 1\}.$$

We will use the following inequalities derived in Serfling (1979). For any probability density g,

$$\int|g(x) - f(x)|dx \le 2C_{f,q}^{1/(q+1)}\sup_x|g(x) - f(x)|^{q/(q+1)} \tag{2.6a}$$

for $0 < q < \infty$, and

$$\int|g(x) - f(x)|dx \le 2C_{f,\infty}\sup_x|g(x) - f(x)|. \tag{2.6b}$$

We now attack again the term $B(f, f_n)$. Putting $g_{n,u}(x) = f(x - c_n u)$ and assuming $\int K(u)du = 1$, we have

$$B(f, f_n) = \int\int\{[g_{n,u}(x) - f(x)]K(u)du\}^2 dx$$

$$\leq \int\int[g_{n,u}(x) - f(x)]^2 dx\, K(u)du.$$

Since $g_{n,u}(x)$, $-\infty < x < \infty$, is a probability density, we have by (2.6a) that

$$B(f, f_n) \leq 2C_{f,q}^{1/(q+1)} \int \sup_x |g_{n,u}(x) - f(x)|^{(2q+1)/(q+1)} du.$$

Now assuming that f is Lipschitz of order β on $(-\infty, \infty)$, we have $\sup_x |g_{n,u}(x) - f(x)| \leq A_f\, c_n^{\beta}|z|^{\beta}$ for some constant A_f, and thus

$$B(f, f_n) = O(c_n^{\beta(2q+1)/(q+1)}), \tag{2.7}$$

provided that $C_{f,q} < \infty$ and $\int|u|^{\beta(2q+1)/(q+1)}|K(u)|du < \infty$. Likewise, if $C_{f,\infty} < \infty$, we obtain $O(c_n^{2\beta})$ in (2.7). Thus we have

THEOREM 2. *Let* f *be Lipschitz of order* β *on* $(-\infty, \infty)$ *and let* $C_{f,q} < \infty$, *where* $0 < q \leq \infty$. *Let* K *satisfy* $\int K(z)dz = 1$ *and* $\int |z|^{\beta(2q+1)/(q+1)}|K(z)|dz < \infty$. *Take* $c_n = Cn^{-(q+1)/[(2\beta+1)q+1]}$. *Then*

$$J(f, f_n) = O(n^{-\beta(2q+1)/[(2\beta+1)q+1]}). \tag{E}$$

In particular, for f Lipschitz of order 1 on $(-\infty, \infty)$ and having bounded support, $J(f, f_n) = O(n^{-2/3})$, the same as in Theorem 1. For $\beta < 1$ or $q < \infty$, however, the rate in (E) is slower than $O(n^{-2/3})$.

Conditions (A) - (E) comprise a hierarchy of possible rates for $J(f, f_n)$. Rates close to $O(n^{-1})$ are achieved in (A), in (B) for p suffiently large, and in (C) for s sufficiently large. On the other hand, the rates in (C) and (D) are no slower than $O(n^{-4/5})$ and $O(n^{-2/3})$, respectively. The rates in (E) may be as high as $O(n^{-2/3})$ but may be as slow as $O(n^{-\gamma})$, any $\gamma > 0$, as may be the rates in (B) for p sufficiently close to $\frac{1}{2}$.

Some interrelationships among conditions on the characteristic function of f, conditions on the derivatives of f, and conditions on the tails of f are as follows. If f has an integrable s-th derivative, then Φ_f decreases faster than an algebraic rate $p > s$ (Feller (1966), p. 487). On the other hand, it is readily proved that if Φ_f decreases at algebraic rate of degree $p > s$, s an integer, then f has a bounded continuous derivative of order $s - 1$. However, conditions on the *tails* of f relate to properties of Φ_f at the *origin*. Namely, if f has q-th absolute moment, q an integer, then Φ_f has a q-th derivative,; conversely, if Φ has a q-th order derivative at the origin, then f has all moments up to order q if q even, $q - 1$ if q odd.

Finally, we remark that in view of the lemma of Section 1 the $O(\cdot)$ rates specified in (A) - (E) apply equally well to $J(f, \tilde{f}_n)$, where $\tilde{f}_n$ is the recursive counterpart of f_n.

REFERENCES

[1] Davies, H. I. and Wegman, E. J. (1975). Sequential nonparametric density estimation. *IEEE Trans. Info. Th.* IT-21, 619-628.

[2] Feller, W. (1966). *An Introduction to Probability Theory and Its applications*, Vol. II. Wiley, New York.

[3] Fryer, M. J. (1977). A review of some non-parametric methods of density estimation. *J. Inst. Maths Applics* 20, 335-354.

[4] Leadbetter, M. R. (1963). On the Non-Parametric Estimation of Probability Densities. Research Triangle Institute Technical Report No. 11 (Doctoral dissertation, University of North Carolina).

[5] Nadaraya, E. A. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theor. Prob. Prob. Applic.* 19, 133-141.

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

[7] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27, 832-837.

[8] Serfling, R. J. (1979). A Variation on Scheffé's Theorem, with Application to Nonparametric Density Estimation. Florida State University Statisstics Report M502 (ONR Technical Report No. 137).

[9] Tapia, R. A. and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*. The Johns Hopkins University Press, Baltimore.

[10] Watson, G. S. and Leadbetter, M. R. (1963). On the estimation of the probability density, I. *Ann. Math. Statist.* 34, 480-491.

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

[12] Wegman, E. J. and Davies, H. I. (1979). Remarks on some recursive estimators of a probability density. *Ann. Statist.* 7, 316-327.

[13] Whittle, P. (1958). On the smoothing of probability density functions. *JRSS B* 20, 334-343.

[14] Winter B. B. (1978). Recursive Nonparametric Estimation of Densities and Some Density Functionals. Department of Mathematics, University of Ottowa.

[15] Wolverton, C. T. and Wagner, T. J. (1969). Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Info. Th.* IT-15, 258-265.

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

UNCLASSIFIED

Security Classification of this Page

REPORT DOCUMENTATION PAGE

1. REPORT NUMBERS FSU No. M505 ONR No. 138	2. GOVT. ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE Convergence Rates for the Mean Integrated Squared Errors of Some Nonparametric Density Estimators of Recursive δ-Function Type		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M505
7. AUTHOR(s) R. J. Serfling		8. CONTRACT OR GRANT NUMBER(s) ONR No. N00014-76-C-0608
9. PERFORMING ORGANIZATION NAME & ADDRESS The Florida State University Department of Statistics Tallahassee, Florida 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA AND WORK UNIT NOS.
11. CONTROLLING OFFICE NAME & ADDRESS Office of Naval Research Statistics & Probability Program Arlington, Virginia 22217		12. REPORT DATE July, 1979 13. NUMBER OF PAGES 11
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Nonparametric density estimation; mean integrated squared error; convergence rates; recursive estimators.

20. ABSTRACT

For estimation of a probability density function f by an empirical function f_n (sub n) based on a sample of size n from f, a widely used measure of goodness is the mean integrated squared error. For the well known δ (delta)-function type of f_n (sub n), we show that the asymptotic behavior of this measure is essentially unchanged if f_n (sub n) is replaced by a recursive version. Also, we characterize this asymptotic behavior under somewhat milder smoothness restrictions on f than previously considered in the literature, at the expense however of adding tail restrictions on f.