King Fahd University of Petroleum & Minerals

DEPARTMENT OF MATHEMATICAL SCIENCES

Technical Report Series

TR 012

November 1980

A Note on Nonparametric Density Estimation for Dependent Variables Using Delta Sequence

Abrahim A. Ahmad

DHAHRAN 31261 ● SAUDI ARABIA ● www.kfupm.edu.sa/math/ ● E-mail: mathdept@kfupm.edu.sa

A NOTE ON NONPARAMETRIC DENSITY ESTIMATION FOR DEPENDENT VARIABLES USING DELTA SEQUENCE

By

Ibrahim A. Ahmad
Department of Mathematical Sciences
University of Petroleum & Minerals
Dhahran, Saudi Arabia

ABSTRACT

A general method based on "delta sequences" due to Walter and Blum (1979, Ann. Statist., 7, 328-340) is extended to sequences of strictly stationary mixing random variables having the same marginal distribution admitting a Lebeseque probability density function. It is proved that, under certain conditions, the rate of mean square convergence obtained in the iid case by Walter and Blum, continues to hold.

AMS Subject Classification: Primary 62G05.
Key words and phrases: Density estimation, delta sequences, mean square, rates of convergence, mixing sequences, stationarity.

1. Introduction and Definitions.

Recently, Walter and Blum (1979) proposed a method for density estimation based on a random sample of n observations, which is more general than most previously known techniques and they established a rate of mean square convergence for densities in Sololev spaces and for densities satisfying Lipschitz conditions.

In the literature there are several types of density estimates and for surveys of the subject we refer the reader to Wegman (1972), Frayer (1977), and Wretz and Schneider (1979). While most papers written on density estimation deal with the case of independent identically distributed (iid) random variables, some of the most recent work is devoted to the case when sampling from stationary sequence of dependent variables, see Bosq (1973), Borwanker (1971), and Ahmad and Lin (1976) among others.

In this note we extend the method of delta-sequences of Walter and Blum (1979) and obtain estimates of the marginal densities of strictly stationary mixing processes and study the rate of mean square convergence of these estimates.

Definition 1.1. Let $\{X_n\}$ *be a sequence of random variables and let* $G(m,n)$ *denote the* σ*-field generated by* $X_{m+1},\dots,X_{m+n}$ *for all* $0 \le m < n \le \infty$*. Further let* $A \in G(1,m)$ *and* $B \in G(m+n,\infty)$ *and define integer-valued functions* $\phi_i(\cdot)$, $i=1,2,3$ *such that* $\phi_i(n) \downarrow 0$ *as* $n \to \infty$, $i=1,2,3$. *Then* $\{X_n\}$ *is said to be:*

(i) ϕ_1*-mixing if* $|P(AB)-P(A)P(B)| \le \phi_1(n)$,

(ii) ϕ_2*-mixing if* $|P(AB) - P(A)P(B)| \le \phi_2(n)P(A)$, *and*

(iii) ϕ_3-mixing if $|P(AB) - P(AB)-(B)| \leq \phi_3(n)P(A)P(B)$.

Note that ϕ_1-mixing is also known as "strong mixing" and it is due to Rosenblatt (1956), while the ϕ_2-mixing is sometimes called "uniform mixing" and is due to Ibragimov (1962), and finally, ϕ_3-mixing is due to Philip (1969). Note also that, ϕ_1-mixing implies ϕ_2-mixing which in turn implies ϕ_3-mixing but the canverses are not necessarily true. The next definition is due to Walter and Blum (1979).

Definition 1.2. *A sequence $\{\delta_m(x,t)\}$ of bounded measurable functions defined on IXI, where I is an open subset of reals, is said to be* "a delta sequence" *on I if for every $x \in I$ and each C^∞ function ψ with support in I we have*

$$\int_I \delta_m(x,t)\psi(t)dt \to \psi(x) \quad \text{as} \quad m \to \infty. \tag{1.1}$$

Using the first n observations from a sequence of strictly stationary ϕ_i-mixing random variables, define the estimator:

$$\hat{f}_{m,n}(x) = n^{-1}\Sigma_{j=1}^{n}\delta_m(x,X_j). \tag{1.2}$$

Walter and Blum (1979) give several examples demonstrating the generality of (1.2). These include the following:

(a) $\delta_m(x,t) = g_m(x-t)$ with $g_m(y)$ the density of the sample mean, and

(b) $\delta_m(x,t) = m\chi_{m^{-1}}(x-t)$, where $\chi_{m^{-1}}(\cdot)$ is the indicator function of the interval $[0,m^{-1}]$.

Let $W_p^{(s)}(I)$ denote the Sobolev space of functions defined on I whose first $s-1$ derivatives are absolutely continuous and whose s-th derivative is in L_p. In order to obtain rates in $E[\hat{f}_{m,n}(x)-f(x)]^2$, Walter and Blum (1979) restrict their attention to two classes of delta sequences; the class of delta sequences that has a (s,q) rate $m^{-\beta}$ and the class of dominated delta sequences of Fejer type. For the former they obtain uniform rate while for the latter they obtain a pointwise rate. For the sake of completeness we list these two definitions.

Definition 1.3. Let $\{\delta_m(x,t)\}$ be a delta sequence on $I=(a,b)$ such that:

(i) $\delta_m(x,.) \in L^2(a,b)$ for all $x \in I$.

(ii) $\|\delta_m(x,.)\|_2 = O(m^{1/2})$ uniformly in $x \in I$.

Let s be a positive integer; denote by $\delta_m^{(-s)}(x,t)$ the antiderivative of order s with respect to t of $\delta_m(x,t)$ which, together with its first $(s-1)$ derivatives, is zero at a. Denote by $\delta_x^{(-s)}(t)$ the function $(x-t)_+^{s-1}/(s-1)!$. For such an s, suppose that there are numbers $q \geq 1$ and $0 < \beta \leq s+q^{-1}-1$ such that:

(iii) $\delta_m^{(-s)}(x,.) - \delta_x^{(-s)}(\cdot) \in L^q(a,b)$ for all $x \in I$.

(iv) $\|\delta_m^{(-s)}(x,.) - \delta_x^{(-s)}(\cdot)\|_q = O(m^{-\beta})$ uniformly in $x \in I$.

(v) $|\delta_m^{(-k)}(x,b) - \delta_x^{(-k)}(b)| = O(m^{-\beta})$ uniformly in $x \in I$ and for $k=1,\ldots,s$.

Then $\{\delta_m(x,t)\}$ is said to have (s,q) rate $m^{-\beta}$.

Some of the examples cited by Walter and Blum satisfy this definition.

Definition 1.4. A delta sequence $\{\delta_m(x,t)\}$ *is said to be dominated by a kernel* K *if for* $x,t \in R$,

(i) $\delta_m(x,t) \geq 0$, $\int \delta_m(x,t)dt = 1$, *and* $\delta_m(x,t) \leq CmK(m(t-x))$.

(ii) $K(u)$ *is such that* $\int K(u)du = 1$ *and* $\int |u| K(u)du < \infty$ *for* $q=1$ *or* $\int |u|^q \ell n_+ |u| K^q(u)du < \infty$ *for* $q > 1$.

A dominated delta sequence defined on $(-\infty,\infty)$ *is said to be a Fejer type if its dominating kernel is* $[\pi(1+t^2)]^{-1}$.

The purpose of the present note is to obtain conditions under which the Walter and Blum (1979) rates of mean square convergence for $\hat{f}_{m,n}(x)$ continue to hold under each of the above three types of mixing strictly stationary random variables, e.g., we show that the rates obtained by Walter and Blum (1979) continue to hold for ϕ_3-mixing assuming that $\Sigma_{n=1}^{\infty}\phi_3(n)<\infty$, which holds true, e.g., if the variables are taken from a Markov process and hence the results of Walter and Blum (1979) continue to hold for Markov sequences. A bit stronger conditions are required for ϕ_1 and ϕ_2-mixings.

2. Rates of Mean Sqare Error.

THEOREM 2.1. *Let* $s \geq 1$ *and let* $f \in W_p^{(s)}(I)$. *Let* $\{\delta_m\}$ *be a delta sequence of* (s,q) *rate* $m^{-\beta}$ *where* q *satisfies* $p^{-1}+q^{-1} = 1$.

(i) *If* $\{X_n\}$ *is strictly stationary* ϕ_1*-mixing such that* $\Sigma_{n=1}^{\infty}[\phi_1(n)]^{1-r_1^{-1}-r_2^{-1}} < \infty$ *for some* $r_1 > 1$ *and* $r_2 > 1$ *with* $r_1^{-1} + r_2^{-1} < 1$, *then*

$$E[\hat{f}_n(x)-f(x)]^2 = 0(n^{-1+1/(1+2\beta)}), \qquad (2.1)$$

uniformly on I, where $\hat{f}_n$ is defined to be $\hat{f}_{m,n}$ of (1.2) with $m=n^{1/(1+2\beta)}$.

(ii) If $\{X_n\}$ is strictly stationary ϕ_2-mixing such that $\Sigma_{n=1}^{\infty}[\phi_2(n)]^{1-r^{-1}} < \infty$ for some $r > 1$, then (2.1) continues to hold.

(iii) If $\{X_n\}$ is strictly stationary ϕ_3-mixing such that $\Sigma_{n=1}^{\infty}\phi_3(n) < \infty$, then (2.1) continues to hold.

Proof. (i) Note that

$$E[\hat{f}_{m,n}(x)-f(x)]^2 = \text{Var } \hat{f}_{m,n}(x) + [E\hat{f}_{m,n}(x)-f(x)]^2. \tag{2.2}$$

But it follows from (4) of Walter and Blum (1979) and the stationarity of $\{X_n\}$ that

$$[E\hat{f}_{m,n}(x)-f(x)]^2 = [\int_I \delta_m(x,t)f(t)dt - f(t)]^2 = 0(n^{-2\beta}). \tag{2.3}$$

Next, we need to evaluate $\text{Var } \hat{f}_{m,n}(x)$. Again by stationarity of $\{X_n\}$,

$$\text{Var } \hat{f}_{m,n}(x) = n^{-1} \text{ Var } \delta_m(x,X_1) + 2n^{-2}\Sigma_{j=2}^{n}(n-j+1) \text{ Cov}(\delta_m(x,X_1),\delta_m(x,X_j)). \tag{2.4}$$

But again, it follows from (3) of Walter and Blum (1979) that $\text{Var } \delta_m(x,X_1) = 0(m)$, uniformly in $x \in I$. Next, by an application of a Lemma of Deo (1973) we see that for all $j=2,\ldots,n$

$$|\text{Cov}(\delta_m(x,X_1),\delta_m(x,X_j))| \le 10[\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}}[E|\delta_m(x,X_1)|^{r_1}]^{r_1^{-1}}[E|\delta_m(x,X_j)|^{r_2}]^{r_2^{-1}}, \tag{2.5}$$

for some $r_1 > 1$, $r_2 > 1$ such that $r_1^{-1} + r_2^{-1} < 1$. But if $r \in (1,2]$, then

$$E^{1/r}|\delta_m(x,X_1)|^r \le E^{\frac{1}{2}}|\delta_m(x,X_1)|^2 \le [\|f\|_\infty \|\delta_m(x,.)\|_2^2]^{\frac{1}{2}} = 0(m^{\frac{1}{2}}), \tag{2.6}$$

where $\|f\|_\infty < \infty$, since $|f^p(x) - f^p(a)| \le p\int_{a_2}^{x}(f(x))^{p-1}f'(x)|dx \le p\|f\|_p^{(p/q)}.\|f'\|_p$.

While if $r \geq 2$,

$$E^{1/r}|\delta_m(x,X_1)|^r \leq C_1 E^{1/r}|\delta_m(x,X_1)|^2 \leq O(m^{1/r}) \leq O(m^{\frac{1}{2}}),$$

since $\{\delta_m\}$ is a delta sequence. Substituting into the right hand side of (2.6) we get for sufficiently large m and a positive constant C_2 that for m sufficiently large

$$|\mathrm{Cov}(\delta_m(x,X_1),\delta_m(x,X_j))| \leq C_2[\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}} m. \tag{2.7}$$

Hence

$$\begin{aligned}
&|\Sigma_{j=2}^n (n-j+1)\mathrm{Cov}(\delta_m(x,X_1),\delta_m(x,X_j))| \\
&\quad \leq C_3 m \Sigma_{j=2}^n (n-j+1)[\phi_1(j)]^{1-r_1^{-1}-r_2^{-2}} \leq C_3 nm \Sigma_{j=1}^\infty [\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}} \\
&\quad \leq C_4 nm,
\end{aligned} \tag{2.8}$$

since $\Sigma_{j=1}^\infty [\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}} < \infty$. Thus collecting terms we arrive at

$$E[\hat{f}_{m,n}(x)-f(x)]^2 = O(m/n) + O(n^{-2\beta}). \tag{2.9}$$

Choosing $m=n^{1/(1+2\beta)}$, the desired conclusion follows.

(ii) Proceed exactly as in Part (i) except we use Lemma 17.2.3 of Ibragimov and Linnik (1971) in (2.5) above to get

$$|\mathrm{Cov}(\delta_m(x,X_1),\delta_m(x,X_j))| \leq 2[\phi_2(j)]^{1-r_1^{-1}}[E|\delta_m(x,X_1)|^{r_1}]^{r_1^{-1}}[E|\delta_m\ x,X_j)|^{r_2}]^{r_2^{-1}}, \tag{2.10}$$

where $r_2^{-1} = 1 - r_1^{-1}$.

(iii) Again the proof follows that of Part (i) but we put Lemma 3 of Philipp (1969) in use and get

$$|\mathrm{Cov}(\delta_m(x,X_1),\delta_m(x,X_j))| \leq \phi_3(j)E|\delta_m(x,X_1)|E|\delta_m(x,X_j)| \leq \phi_3(j)E|\delta_m(x,X_j)|^2. \tag{2.11}$$

The theorem is now proved. □

Remark 2.2. If the summability conditions of ϕ_i's are difficult to verify, it is possible to reduce them and obtain a somewhat weaker rates of the mean square convergence, precisely, if

(i) $\phi_1(n) = 0(n^{-1/(1-r_1^{-1}-r_2^{-1})})$ for some $r_1, r_2 > 1$ and $r_1^{-1}+r_2^{-1}<1$, or

(ii) $\phi_2(n) = 0(n^{-1/(1-r^{-1})})$ for some $r > 1$, or

(iii) $\phi_3(n) = 0(n^{-1})$, then

$$E[\hat{f}_n(x)-f(x)]^2 = 0(n^{-1+1/(1+2\beta)}\ln n). \tag{2.12}$$

In order to demonstrate that (2.12) is valid we observe from (2.8) that

$$|\Sigma_{j=2}^{n}(n-j+1)\mathrm{Cov}(\delta_m(x,X_1),\delta_m(x,X_j))|$$
$$\leq C_3 mn\Sigma_{j=1}^{n}[\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}} \leq C_4 mn \ln n,$$

for sufficiently large n, since $\phi_1(n) = 0(n^{-1/(1-r_1^{-1}-r_2^{-1})})$. Thus we obtain a bound in (2.9) equal to $0(\ln n)[0(m/n) + 0(m^{-2\beta})]$ from which (2.12) follows with the choice $m=n^{1/(1+2\beta)}$.

Note also that taking $\ln n \leq n^{\gamma}$ for some $\gamma > 0$, then we get under either of the above conditions that

$$E[\hat{f}_n(x)-f(x)]^2 = 0(n^{-1+1/(1+2\beta)+\gamma}). \tag{2.13}$$

Remark 2.3. A delta sequence which is dominated by a kernal K in Definition 1.4 has $(1,q)$ rate $m^{-1/q}$, see Corollary 1 of Walter and Blum (1979), and thus Theorem 2.1 above applies for such sequences.

Next, we obtain a pointwise rate of convergence of the mean square error for the class of dominated delta sequence of Fejer type when $f(x)$ is Lipschitz of order λ, $0 < \lambda < 1$.

THEOREM 2.4. *Let f be a bounded density satisfying Lipschitz condition of order λ, $0 < \lambda < 1$ at $x=0$. Let $\{\delta_m\}$ be a delta sequence of Fejer type.*

(i) *If $\{X_n\}$ is strictly stationary ϕ_1-mixing such that $\Sigma_{n=1}^{\infty} [\phi_1(n)]^{1-r_1^{-1}-r_2^{-1}} < \infty$ for some $r_1 > 1$ and $r_2 > 1$ such that $r_1^{-1} + r_2^{-1} < 1$, then the estimator given by $\hat{f}_n = \hat{f}_{m,n}$ with $m = [n^{1/(1+2\lambda)}]$ and*

$$\hat{f}_{m,n} = \hat{f}_{m,n}(0) = n^{-1} \Sigma_{i=1}^{n} \delta_m(X_i), \tag{2.14}$$

satisfies

$$E[\hat{f}_n - f(0)]^2 = O(n^{-1+1/(1+2\lambda)}). \tag{2.15}$$

(ii) *If $\{X_n\}$ is strictly stationary ϕ_2-mixing such that $\Sigma_{n=1}^{\infty} [\phi_2(n)]^{1-r^{-1}} < \infty$ for some $r > 1$, then the conclusion in (2.15) continues to hold.*

(iii) *If $\{X_n\}$ is strictly stationary ϕ_3-mixing such that $\Sigma_{n=1}^{\infty} \phi_3(n) < \infty$, then the conclusion in (2.15) continues to hold.*

Proof. Again we shall prove Part (i) only, the other two parts may be proven as in Theorem 2.1 above.

(i) Note that

$$E[\hat{f}_n - f(0)]^2 = \text{Var } \hat{f}_n + [E\hat{f}_n - f(0)]^2 \tag{2.16}$$

But by splitting the integral into five parts, Walter and Blum (1979), Theorem 2, show that $[E\hat{f}_n - f(0)]^2 = O(n^{-2\lambda})$, with $m = [n^{1/(1+2\lambda)}]$. Thus we need only to handle the first term in the right-hand-side of (2.16). By stationarity,

$$\operatorname{Var}\, \hat{f}_n = n^{-1}\operatorname{Var}\ \delta_m(X_1) + 2n^{-2}\Sigma_{j=2}^{n}(n-j+1)\operatorname{Cov}(\delta_m(X_1), \delta_m(X_j)). \quad (2.17)$$

But since $\{\delta_m\}$ is a delta sequence of Fejer type, then $n^{-1}\operatorname{Var}\ \delta_m(X_1) \leq Cm/n$ for some constant $C > 0$. Finally using a Lemma of Deo (1973) we get

$$\begin{aligned} &|n^{-2}\Sigma_{j=2}^{n}(n-j+1)\operatorname{Cov}(\delta_m(X_1),\ \delta_m(X_j))| \\ &\quad \leq 10n^{-2}[E|\delta_m(X_1)|^{r_1}]^{r_1^{-1}}[E|\delta_m(X_1)|^{r_2}]^{r_2^{-1}}\Sigma_{j=2}^{n}(n-j+1)[\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}} \\ &\quad \leq 10n^{-1}CE^{1/r_1}|\delta_m(X_1)|^{r_1}E^{1/r_2}|\delta_m(X_1)|^{r_2}, \end{aligned} \quad (2.18)$$

since $\Sigma_{j=2}^{n}[\phi_1(j)]^{1-r_1^{-1}-r_2^{-1}} \leq \Sigma_{n=1}^{\infty}[\phi_1(n)]^{1-r_1^{-1}-r_2^{-1}} < \infty$ and using the stationarity of $\{X_n\}$. But by Definition 1.2, if $r \in (1,2]$,

$$E^{1/r}|\delta_m(X_1)|^r \leq E^{1/2}|\delta_m(X_1)|^2 \leq O(m^{1/2}), \quad (2.19)$$

and if $r > 2$, since $\{\delta_m\}$ is a delta function sequence, then

$$E^{1/r}|\delta_m(X_1)|^r \leq CE^{1/r}|\delta_m(X_1)|^2 \leq O(m^{1/r}) \leq O(m^{1/2}). \quad (2.20)$$

Hence using (2.19) and (2.20) into (2.18) we easily see that

$$n^{-2}\Sigma_{j=2}^{n}(n-j+1)\operatorname{Cov}(\delta_m(X_1), \delta_m(X_j)) = O(m/n). \quad (2.21)$$

Now the desired conclusion follows from the fact that $[E\hat{f}_n - f(0)]^2 = O(m/n) + O(m^{2\lambda})$, with $m = n^{1/(1+2\lambda)}]$. □

Remark 2.5. As pointed out in Remark 2.2, if we reduce the conditions on ϕ_i's we obtain slower rates of $E[\hat{f}_n - f(0)]^2$, in fact if in Part (i) (Part (ii) or Part (iii)) we only assume that $\phi_1(n) = 0(n^{-1(1-r_1^{-1}-r_2^{-1})})$ $(\phi_2 = 0(n^{-1/(1-r^{-1})})$ or $\phi_3(n) = 0(n^{-1}))$, then we obtain the rate

$$E[\hat{f}_n - f(0)]^2 = 0(n^{-1+1/(1+2\lambda)} \ln n). \tag{2.22}$$

Thus taking $\ln \leq n^{\gamma}$ for some $\gamma > 0$ we get

$$E[\hat{f}_n - f(0)]^2 = 0(n^{-1+1/(1+2\lambda)+\gamma}). \tag{2.23}$$

Note also that in the iid case the rate of Theorem 2.1 is the best possible for $f \in W_p^{(n)}(I)$, see Wahba (1975), while the rate in Theorem 2.4 is the best possible for Lipschitz densities, see Farrell (1972), hence our results present conditions under which the best possible rates are obtainable under ϕ_i-mixing conditions, i=1,2,3.

REFERENCES

[1] Ahmad, I.A. and Lin, P.E. (1975).. Nonparametric estimation of probability density functions for dependent variables with applications. Under revision.

[2] Borwanker, J. (1971). Asymptotic theory of density estimation. Z. Wahrscheinlichkeitstheorie verw. Geb., 20, 182-188.

[3] Bosq, D. (1973). Sur l'estimation de la densite d'un processes stationair et melangeant. C.R. Acad. Sci. Paris, T227, 535-538.

[4] Deo, C.M. (1973). A note on empirical processes for strong mixing sequences. Ann. Prob., 1, 870-875.

[5] Farrell, R.H. (1972). On best obtainable asymptotic rates of convergence in estimation of a density function at a point. Ann. Math. Statist., 43, 170-180.

[6] Fryer, M.J. (1977). A review of some nonparametric methods of density estimation. J. Inst. Math. Appl., 20, 335-354.

[7] Ibragimov, I.A. (1962). Some limit theorems for stationary sequences. Theor. Probability Appl., 7, 349-382.

[8] Ibragimov, I.A. and Linnik, Yu.V. (1971). Independent and Stationary Sequences of Random Variables. Walters-Noordhoff, Netherlands.

[9] Philipp, W. (1969). The central limit theorem for mixing sequences of random variables. Z. Wahrscheinlickeitstheorie verw Geb., 12, 155-171.

[10] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. USA, 42, 43-47.

[11] Wahba, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. Ann. Statist. 3, 15-29.

[12] Walter, G. and Blum, J. (1979). Probability density estimation using delta-sequences. Ann. Statist., 7, 328-340.

[13] Wegman, E.J. (1972). Nonparametric probability density estimation: I. A summary of available methods. Technometrics,14, 316-327.

[14] Wretz, W. and Schneider, B. (1979). Statistical density estimation: A bibloigraphy. Int. Statist. Review, 47, 155-175.