Journal of the Korean Statistical Society (1998), **27**: *2, pp 153–163*

Kernel Density Estimation in the L^∞ Norm under Dependence †

Tae Yoon Kim[1]

ABSTRACT

We investigate density estimation problem in the L^∞ norm and show that the iid optimal minimax rates are achieved for smooth classes of weakly dependent stationary sequences. Our results are then applied to give uniform convergence rates for various problems including the Gibbs sampler.

Keywords: Kernel density estimation; Uniform convergence; Optimal rate; Mixing

1. INTRODUCTION

Recently research interest has been surging in density estimation problems under dependence. For example, estimation for the stationary p.d.f. often might be required for a Markov chain, an ARMA sequence, or the Gibbs sampler. This article considers kernel density estimation problems under dependence, particularly the iid optimal minimax rates in L^∞ norm. Tran (1990) obtained some rates for weakly dependent variables under the additional condition that smoothing parameter h_n tends to zero more slowly than the iid case. See, e.g., Theorem 2.1 of Tran. Yu (1993) got the iid optimal minimax rates for β-mixing under a similar condition on h_n (see remark 2.2). These assumed conditions on h_n seem to be unattractive because they usually depend on the unknown parameters. The purpose of this article is to show that the iid optimal rates for α-mixing hold without referring to the unattractive conditions on h_n. Our result answers the question raised by Yu whether the iid optimal rates continue to hold for α-mixing sequences. Note that α-mixing is weaker than β-mixing. Application of our results to some interesting problems will be given in section 2.

We briefly discuss the mathematical framework of estimating stationary density and its derivatives by kernel density estimator for a stationary sequence. Let

†This paper was supported by NON DIRECTED FUND, Korea Research Foundation, 1996.
[1]Department of Mathematics and Statistics, Keimyung University, Taegu 704-701, Korea.

$\mathbf{X} = (X_1, \dots, X_n, \dots)$ be a sequence of random variables with domain $\mathcal{D}$ in R^d. $\alpha(n)$ is called the α- mixing coefficient if

$$\alpha(n) = \sup_k \sup \{|P(A \cap B) - P(A)P(B)| : A \in \sigma(X_1, \dots, X_k), \\ B \in \sigma(X_{k+n}, \dots)\}.$$

Suppose $X_1, \dots, X_n$ is a fragment of an α-mixing sequence of r.v.'s with a density function on $\mathcal{D}$. We would like to use a kernel estimator to estimate not only the density but also its derivatives. Let $\alpha = (\alpha_1, \dots, \alpha_d)$ denote a nonnegative integer $d-$tuple, $[\alpha] = \alpha_1 + \cdots + \alpha_d$, $\alpha! = \alpha_1! \cdots \alpha_d!$. Then for any $x \in R^d$, denote $D^\alpha = \partial^{[\alpha]}/\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}$. Let $Q = \sum_{[\alpha] \le m} q_\alpha D^\alpha$, where the q's are real constants and $q_\alpha \ne 0$ for some $[\alpha] \le m$. Suppose we are interested in estimating $Qf(x) = \sum_{[\alpha] \le m} q_\alpha D^\alpha f(x)$. Without loss of generality, we may assume $Q = D^m$; hence based on a stationary mixing sequence $X_1, \dots, X_n$, the kernel estimator with a bandwidth $0 < h_n < 1$ is

$$\hat{f}_n(x) = (nh_n^d)^{-1} \sum_{i=1}^n K(\frac{X_i - x}{h_n})$$

and $D^m f$ can be estimated by

$$D^m \hat{f}(x) = n^{-1} h_n^{-d-m} \sum_{i=1}^n (D^m K)(\frac{X_i - x}{h_n}),$$

where $K(\cdot)$ is a bounded kernel on R^d.

2. MAIN RESULTS AND ITS APPLICATION

In this section, we give the main results on the kernel density estimation in the L^∞ norm for stationary α-mixing sequences. For simplicity, we introduce a bias assumption and an algebraic growth condition used by Yu (1993).

Bias assumption of order (p, m). Let $\mathcal{F}$ be a class of densities. If there exists an $C_\mathcal{F} > 0$ such that for any $0 < h_n \to 0$ and the kernel estimator $\hat{f}_n$ with the bandwidth h_n

$$\sup_{f \in \mathcal{F}} |ED^m \hat{f}_n(x) - D^m f(x)| \le C_\mathcal{F} h_n^{p-m},$$

then we say that $(\mathcal{F}, K)$ satisfies the bias assumption of order (p, m).

Algebraic growth condition. Let

$$G_n(K,m) = \{(D^m K)(h_n^{-1}(x-\cdot)) : x \in \mathcal{D}\}$$

for a particular sequence $h_n \to 0$. For the class G_n of functions and a sequence $\epsilon_n \downarrow 0$, we say (G_n, ϵ_n) satisfies the algebraic growth condition if for some positive constants C and w

$$\sup_\mu N_1(\epsilon_n, \mu, G_n) \le Cn^w,$$

where N_1 is L^1 covering number defined as:

$$N_1(\epsilon,\mu,G_n) = \min\Big\{k : g_1,\dots,g_k \in L^1(\mu) \text{ such that}$$
$$\min_{1\le j\le k} \int |g(x)-g_j(x)|d\mu(x) \le \epsilon, g \in G_n\Big\} \tag{2.1}$$

for any distribution μ on $\mathcal{D}$ and a class G_n of functions in $L^1(\mu)$.

Theorem 2.1. *Suppose* $\mathbf{X}$ *is a stationary* α*-mixing sequence with the mixing coefficient* $\alpha(n) = n^{-\kappa}$ *for some* $\kappa > 0$. *Let* $h_n = (n^{-1}\log n)^{1/(2p+d)}$. *Assume that* $(\mathcal{F}, K)$ *satisfies the bias assumption of order* (p,m) *and that* g*'s found in definition (2.1) are independent of probability measure* μ *and hence* $G_n = G_n(K,m)$ *satisfies the algebraic growth condition with an exponent* $w > 0$. *If*

$$E(g(X_i)g(X_j)) \le Ch_n^{2d} \text{ for any } g \in G_n \text{ and } i \ne j \tag{2.2}$$

and

$$\kappa > \{2(w+1)(2p+d) + 3(d+p)\}/2p, \tag{2.3}$$

then we have

$$\sup_{\mathcal{F}} \sup_{x\in R^d} |D^m\hat{f}(x) - D^m f(x)| = O([n^{-1}\log n]^{(p-m)/(2p+d)}) \text{ a.s.} \tag{2.4}$$

Remark 2.1: It should be observed that the rate of convergence established in Theorem 2.1 is in fact optimal because a sequence of independent random variables is also a stationary sequence. See Stone (1983). Note that the bias assumption and the algebraic growth condition of Theorem can be shown to hold if $\mathcal{D} = [0,1]^d$, $(D^m)K$ is Hölder continuous, and f belongs to some smooth class of densities. See section 2 of Yu for detailed discussions. Note that Tran (1990) assumed $p = 1$, the Hölder continuity of K, the compactness of domain, and (2.2). In addition (2.2) holds if the bivariate density $f(X_i, X_j)$ is bounded.

Remark 2.2: Recent results have been established under the additional condition that h_n tends to zero more slowly than the iid case. For example Yu (1993) and Tran (1990) have obtained their results under this. In particular Yu's result for β-mixing holds under $2p > d$. Note that $2p > d$ is equivalent to imposing an slower speed of h_n. She removed this at the cost of the compact domain, but it appears to have severe technical problems (see Kim (1998)).

Remark 2.3: Since (2.3) holds for any p, w, and d when the mixing coefficient is geometric mixing (i.e. $\alpha(m) = O(\rho^m)$ for $0 < \rho < 1$), it can be said that the iid optimal rates hold without any additional condition under geometric mixing. It is an useful application of our Theorem, considering that all the examples discussed below are in fact geometric mixing.

Now we use Theorem 2.1 to give the rates of convergence of various density estimators we could often face under dependence. Observe that our results below holds without any additional condition on h_n.

Let $\mathbf{X}$ be a stationary Markov chain. Since Markov chain can be shown to be geometric strong mixing, a direct application of our result yield

$$\sup_x |D^m \hat{f}(x) - D^m f(x)| = O([n^{-1}\log n]^{(p-m)/(2p+d)}) \text{ a.s.} \tag{2.5}$$

$$\sup_{x\in E}\sup_{y\in[0,1]^d} |\hat{f}(y|x) - f(y|x)| = O([n^{-1}\log n]^{p/(2p+2d)}) \text{ a.s.} \tag{2.6}$$

under the conditions of Theorem 2.1. Observe that (2.5) and (2.6) addresses convergence rates for the stationary density and the transition density of the Markov chain respectively. For (2.6), we consider $((X_1, X_2), (X_2, X_3), \dots)$ as a Markov chain in R^{2d} and hence apply our result with d replaced by $2d$. In particular E denotes a compact set on which $f(x) \geq c_0 > 0$ and $\hat{f}(y|x) = \hat{f}(x,y)/\hat{f}(x)$, where $\hat{f}(x,y)$ is a kernel density estimator with another kernel K_1 on R^{2d} with the bandwidth $h_n = O((n^{-1}logn)^{1/(2p+2d)})$.

Let $\{Y(t)\}_{t\in Z}$ be the unique sequence satisfying the following ARMA equation:

$$\sum_{i=1}^{P} B(i)Y(t-i) = \sum_{k=0}^{Q} A(k)\epsilon(t-k),$$

where $B(i)$ and $A(k)$ are dxd and dxr matrices, $B(0) = Id$, $\epsilon(t)$ are iid in R^d, $E\epsilon(t) = 0$. Since by Mokkadam (1988), $Y(t)$ was shown to be geometric mixing, an immediate application of our result yields

$$\sup_y |\hat{f}(y) - f(y)| = O([n^{-1}\log n]^{p/(2p+d)}). \text{ a.s.,} \tag{2.7}$$

where $f(y)$ is the p.d.f. of Y_1.

The Gibbs sampler or its analogy is now a popular computer simulation method to obtain samples from distributions which cannot be sampled from otherwise. See Geman and Geman (1984). In the below, we apply Theorem 2.1 to the Gibbs sampler related convergence rate problems. See also section 4 of Yu. Consider the Gibbs sampler in the case of two variables (X, Y). It is assumed that it is easy to sample from the two conditional probability distributions $f(x|y)$ and $g(x|y)$. To obtain one of the marginal distributions, say $f(x)$, the Gibbs sampler simulates a joint Markov chain whose limiting stationary marginals (equilibrium) are $f(x)$ and $g(y)$; namely, it simulates a joint Markov chain $X_0, Y_0, X_1, Y_1 \dots, X_i, Y_i$ starting with some initial distribution f_0 from which we can sample, then continue by drawing Y_i from $g(\cdot|X_i)$ and X_{i+1} from $f(\cdot|Y_i)$. Now we can use the kernel density estimator to estimate the density function $f(x)$ based on successive samples from the Markov chain. [cf. Liu et al. (1991)]. Using the asymptotic stationarity condition to handle the nonstationarity of the Markov chain generated from Gibbs sampler, one may show that under a set of regular conditions,

$$\sup_x |\hat{f}(x) - f(x)| = O([n^{-1} \log n]^{p/(2p+d)}). \text{ a.s.} \tag{2.8}$$

For details of the regular conditions, refer the condition of Theorem 4.2 in Yu.

3. PROOF

Throughout this section, let M_k denote a bound on both K itself and its derivatives of order greater than m and c denote a generic positive constant. First we provide some lemmas which will be used later.

Lemma 3.1. *If* $\mathbf{X}$ *is* α*-mixng, then for* $p, q, r > 1$*, satisfying* $1/p+1/q+1/r = 1$*, we have*

$$|EX_0X_j - EX_0EX_j| \le 15\alpha^{1/r}(j)\{E|X_0 - EX_0|^p\}^{1/p}\{E|X_j - EX_0|^q\}^{1/q}.$$

Proof: See Theorem 17.2.1 of Ibragimov and Linnik (1971). □

Observing that for $g \in G_n(k, m)$ and $f \in \mathcal{F}$, $|g| \le M_k$, $|f| \le M$ and $|Eg(X_1)| = |E(D^m K)((x - X)/h_n)| \le ch_n^d$, it is easy to see that, for $r \ge 1$

$$E|g(X_1) - Eg(X_1)|^r \le ch_n^d. \tag{3.1}$$

Lemma 3.2. *Let* $V = \sum_{i=1}^{b_n}(g(X_i) - E(g(X_i))$ *for* $b_n \le n$ *and assume (2.2). If* $\mathbf{X}$ *is* α*-mixng satisfying (2.3), then*

$$E(V)^2 \le cb_n h_n^d$$

uniformly over $g \in G(K, m)$.

Proof: Denote $\nu = Eg(X_1)$.

$$\begin{aligned} E(V)^2 &= E[\sum_{i=1}^{b_n}(g(X_i) - Eg(X_i))]^2 \\ &= \sum_{i=1}^{b_n} E(g(X_i) - \nu)^2 + \sum_{i \ne k} E(g(X_i) - \nu)(g(X_k) - \nu) \\ &= b_n E(g(X_i) - \nu)^2 + 2\sum_i \sum_k E(g(X_i) - \nu)(g(X_{i+k}) - \nu). \end{aligned}$$

Using (2.2) and (3.1),

$$E(g(X_i) - \nu)(g(X_{i+k}) - \nu) \le \begin{cases} ch_n^{2d} & \text{for } k > 0, \\ ch_n^d & \text{for } k = 0. \end{cases}$$

Since X_i's are α-mixing, by Lemma 3.1,

$$|E(g(X_i) - \nu)(g(X_{i+k}) - \nu)| \le c\alpha(k).$$

Thus

$$E(V)^2 = O\left(b_n h_n^d + b_n \sum_{k=1}^{b_n} \min(\alpha(k), h_n^{2d})\right). \tag{3.2}$$

If $b_n h_n^d < 1$, the right-hand side of (3.2) is bounded by $c(b_n h_n^d + (b_n h_n^d)^2) \le cb_n h_n^d$. If $b_n h_n^d \ge 1$, the right-hand side of (3.2) is bounded by

$$c(b_n h_n^d + b_n q h_n^{2d} + b_n q^{-\kappa+1})$$

with $q = 1/h_n^d \le b_n$. Since $\kappa > 2$ by (2.3), the last expression is less than $cb_n h_n^d$. The desired result follows. □

The following result will be needed to approximate α-mixing r.v.'s by independent ones.

Lemma 3.3. *Suppose $m = 2st$ for some positive integer t. Suppose V_j, $1 \leq j \leq t$ is a sequence of r.v.'s with V_j being measurable with respect to the σ-field $\sigma(X_{(2j-1)s+1}, \cdots, X_{2js})$. Let ξ and γ be positive numbers such that $\xi \leq \| V_j \|_\gamma < \infty$ for all $1 \leq j \leq t$. Then there exists a positive constant c and a sequence of independent r.v.'s W_j, $1 \leq j \leq t$ such that W_j has the same distribution as V_j and satisfies*

$$P[|V_j - W_j| > \xi] \leq c(\| V_j \|_\gamma /\xi)^\tau \alpha(s)^{2\tau} \tag{3.3}$$

where $\tau = \gamma/(2\gamma + 1)$.

The proof of Lemma is given in Bradley (1983).

Proof of Theorem 2.1: a) By the bias assumption of order (p, m), we have

$$\begin{aligned} & \| Q\hat{f}_n(x) - Qf(x) \| \\ \leq\ & \| Q\hat{f}_n(x) - EQ\hat{f}_n(x) \| + \| EQ\hat{f}_n(x) - Qf(x) \| \\ =\ & h_n^{-d-m} \sup_{g \in G_n(K,m)} \left| n^{-1} \sum_{i=1}^{n} g(X_i) - Eg(X_1) \right| + \| EQ\hat{f}_n(x) - Qf(x) \| \\ =\ & O_p(h_n^{-d-m}\epsilon_n) + ch_n^{p-m}, \end{aligned}$$

where $\epsilon_n = \delta[n^{-1}\log n]^{1/2} h_n^{d/2}$ for $\delta > 0$ by (3.4). Then by choosing $h_n = (n^{-1}\log n)^{1/(d+2p)}$, the desired result follows.

Thus our proof will be finished if we show that

$$P\left(\sup_{g \in G_n} \left| n^{-1} \sum_{i=1}^{n} g(X_i) - Eg(X_i) \right| \geq \epsilon_n \right) \leq cn^{-a} \tag{3.4}$$

where $\epsilon_n = \delta[n^{-1}\log n]^{1/2} h_n^{d/2}$ with $\delta > 0$ and $a > 1$. To verify (3.4), we will employ an approximation of weakly dependent r.v.'s by independent r.v.'s. Let

$$\lambda = (nh_n^d)^{-1/2}(\log n)^{1/2} \text{ and } s_n = [(nh_n^d)^{1/2}/(2M_k(\log n)^{1/2})]. \tag{3.5}$$

For a pair of the integers (s_n, t_n) such that $(n - 2s_n) \leq 2s_n t_n \leq n$, we divide the segment of $X_1, \ldots, X_n$ of the mixing sequence into $2t_n$ blocks of size s_n and a remaining block. Then write

$$\sum_{i=1}^{n} (g(X_i) - Eg(X_i)) = S_{1n} + S_{2n} + Re,$$

where

$$S_{1n} = \sum_{j=1}^{t_n} V(n,2j), \quad S_{2n} = \sum_{j=1}^{t_n} V(n,2j-1),$$

and

$$V(n,j) = \sum_{i=(j-1)s_n+1}^{js_n} (g(X_i) - Eg(X_i)), \text{ for } 1 \leq j \leq 2t_n,$$

and

$$Re = \sum_{i=2s_nt_n+1}^{n} (g(X_i) - Eg(X_i)).$$

It is easy to note that $|Re|/n$ can be made smaller than $\epsilon_n/3$ since $s_n = o(n\epsilon_n)$ and g is bounded. Now we will find an appropriate bound for

$$P\left(|S_{1n}| \geq \delta(nh_n^d \log n)^{1/2}/3\right) + P\left(|S_{2n}| \geq \delta(nh_n^d \log n)^{1/2}/3\right).$$

We will refer to $V(n,2j)$ simply as V_j for simplicity. We have

$$V_j = \sum_{(2j-1)s_n+1}^{2js_n} (g(X_i) - Eg(X_i)).$$

Notice that $2s_nt_n$ is an integer. By Lemma 3.3, there exists a sequence of independent r.v.'s W_j, $1 \leq j \leq t_n$ such that W_j has the same distribution as V_j and satisfies (3.3). Now,

$$P\left(|S_{1n}| \geq \delta(nh_n^d \log n)^{1/2}/3\right)$$

$$\leq P\left(\left|\sum_{j=1}^{t_n} W_j\right| \geq \delta(nh_n^d \log n)^{1/2}/6\right) + P\left(\left|\sum_{j=1}^{t_n}(V_j - W_j)\right| \geq \delta(nh_n^d \log n)^{1/2}/6\right).$$

Clearly $\lambda|W_j| \leq \lambda s_n M_k \leq 1/2$ and

$$\exp(\lambda W_j) \leq 1 + \lambda W_j + (\lambda W_j)^2. \tag{3.6}$$

A simple computation shows that

$$\lambda^2(nh_n^d) = \log n \tag{3.7}$$

By Lemma 3.2,

$$\sum_{j=1}^{t_n} EW_j^2 = \sum_{j=1}^{t_n} EV_j^2 \leq ct_ns_nh_n^d \leq cnh_n^d. \tag{3.8}$$

Observe that

$$P\left(\left|\sum_{j=1}^{t_n} W_j\right| \geq \delta(nh_n^d \log n)^{1/2}/6\right)$$

$$\leq P\left(\sum_{j=1}^{t_n} W_j \geq \delta(nh_n^d \log n)^{1/2}/6\right) + P\left(\sum_{j=1}^{t_n} W_j \leq -\delta(nh_n^d \log n)^{1/2}/6\right). \quad (3.9)$$

Let l be an arbitrary large positive number. We first show that

$$P\left(\sum_{j=1}^{t_n} W_j \geq \delta(nh_n^d \log n)^{1/2}/6\right) \leq n^{-l}$$

Using the independence of the W_j's, Markov inequality and (3.6)-(3.8), we have for sufficiently large δ

$$\begin{aligned} P\left(\sum_{j=1}^{t_n} W_j \geq \delta(nh_n^d \log n)^{1/2}/6\right) &\leq \exp\left(-\lambda\delta(nh_n^d \log n)^{1/2}/6 + \lambda^2 \sum_{j=1}^{t_n} EW_j^2\right) \\ &\leq \exp\left((-\delta/6 + c)\log n\right) \leq n^{-l}. \end{aligned} \quad (3.10)$$

In a similar fashion, one may easily show that the second term of (3.9) is bounded by the right-hand side of (3.10). Thus we have

$$P\left(\left|\sum_{j=1}^{t_n} W_j\right| \geq \delta(nh_n^d \log n)^{1/2}/6\right) \leq n^{-l}. \quad (3.11)$$

Next,

$$P\left[|\sum_{j=1}^{t_n}(V_j - W_j)| > \delta(nh_n^d \log n)^{1/2}/6\right]$$

$$\leq t_n \max_{1\leq j\leq t_n} P\left[|(V_j - W_j)| > \delta(nh_n^d \log n)^{1/2}/(6t_n)\right]. \quad (3.12)$$

If $\| V_j \|_\gamma \geq \delta(nh_n^d \log n)^{1/2}/(6t_n)$, then for some positive constant c,

$$\begin{aligned} &P\left[|\sum_{j=1}^{t_n}(V_j - W_j)| > \delta(nh_n^d \log n)^{1/2}/6\right] \\ \leq\ & ct_n[t_n/\log n]^\tau[nh_n^d/\log n]^{-\tau\kappa} \\ =\ & cn^{\{(p+d)(1+\tau)-2p\kappa\tau\}/(2p+d)}(\log n)^{[p-\tau d-\tau p+2p\kappa\tau]/[(2p+d)]}, \end{aligned} \quad (3.13)$$

by (3.3) with $\tau = \gamma/(2\gamma+1)$, (3.5) and (3.12), and $h_n = (n^{-1}\log n)^{1/(d+2p)}$. In the last expression , we used the following, i.e.,

$$\| V_j \|_\gamma \le cs_n \text{ for all } 1 \le j \le t_n \text{ and any } \gamma \ge 1. \tag{3.14}$$

If $\| V_j \|_\gamma \le \delta(nh_n^d \log n)^{1/2}/(6t_n)$, then

$$\begin{aligned} P\left[|\sum_{j=1}^{t_n}(V_j - W_j)| > \delta(nh_n^d \log n)^{1/2}/6 \right] &\le t_n \max_{1\le j\le t_n} P[|V_j - W_j| > \| V_j \|_\gamma] \\ &\le t_n[nh_n^d/\log n]^{-\tau\kappa}, \end{aligned}$$

which is again bounded by (3.13) for large n since $t_n/\log n \to \infty$ as $n \to \infty$.

By the given algebraic growth condition (i.e., the g's are fixed), we have

$$\begin{aligned} &P\left(\sup_{g\in G_n}\left|n^{-1}\sum_{i=1}^{n} g(X_i) - Eg(X_i)\right| \ge \epsilon_n\right) \\ &\le cn^w P\left(\left|n^{-1}\sum_{i=1}^{n} g(X_i) - Eg(X_i)\right| \ge \epsilon_n\right) \end{aligned}$$

for some $w > 0$. Then by (3.11) and (3.13), the last expression is bounded by

$$\le cn^{w-l} + cn^{w+\{(d+p)(1+\tau)-2p\kappa\tau\}/(2p+d)}(\log n)^{[p-\tau d-\tau p+2p\kappa\tau]/(2p+d)}.$$

Thus $w-l < -1$ for sufficiently large l, and $w+\{(p+d)(1+\tau)-2p\kappa\tau\}/(2p+d) < -1$ if

$$\{(w+1)(2p+d) + (d+p)(1+\tau)\}/(2p\tau) < \kappa.$$

Using continuity argument, one may replace $\tau = 1/2$ in the above. Note that one may let $\gamma \to \infty$ by virtue of (3.14). Now we have

$$\{2(w+1)(2p+d) + 3(d+p)\}/2p < \kappa.$$

Thus if (2.3) holds, (3.4) follows from Borel-Cantelli lemma.

REFERENCES

Bradley, R. C.(1983). "Approximation theorems for strongly mixing random variables," *Michigan Math. J.* **30** 69-81.

Geman, S. and Geman, D.(1984). "Stochastic relaxation, Gibbs distributions, and the Baysean restoration of image," *IEEE Trans. Pattern Analysis and Machine Intelligence* **6** 721-741.

Ibragimov, I.A. and Linnik, Y.A.(1971). *Independent and Stationary Sequences of Random Variables.* Wolters Noordohoff, Groningen.

Kim, T.Y.(1998). "A remark on optimal rates for kernel density estimation in the L^{∞} norm under mixing," To appear in *Proceedings of Mathematical Sciences*, Institute of Mathecal Sciences, Keimyung University.

Liu, J., Wong, W., and Kong, A.(1991). "Correlation structure and convergence rate of the Gibbs sampler with various scans," *Technical Report 304* Dept. Statistics, Univ. Chicago.

Mokkadem, A.(1988). "Mixing properties of ARMA processes," *Stochastic Process. Appl.* **29** 309-315

Stone, C. J.(1983). "Optimal uniform rate of convergence for nonparametric estimators of density function or its derivatives," In *Recent Advances in Statistics.* (M. H.Rizvi, J.S. Rustagi and D. Siegmund, eds.) pp. 393-406, Academic, New York.

Tran, L. T.(1990). "Kernel density estimation under dependence," *Statist. Probab. Lett.* **10** 193-201

Yu, B.(1993). "Density estimation in the L^{∞} norm for dependent data with applications to the Gibbs sampler," *Ann. Statist.* **21** 711-735.