Journal of the Korean Statistical Society (2002), **31**: *3, pp 359–368*

Kernel Regression Estimation Under Dependence†

Tae Yoon Kim[1] and Donghoh Kim[2]

Abstract

Nonparametric kernel regression problem is considered for a stationary dependent sequence $\{(X_j, Y_j) \mid j \geq 1\}$. In particular consistency and rates of convergence are discussed, which gives some useful insight for the effect of dependence for stationary α-mixing sequences.

Keywords. Kernel regression estimates, stationarity, convergence rates.
AMS 2000 subject classifications. Primary 62G20, 62E20; Secondary 60F05.

1. Introduction

The focus of this paper is to study nonparametric regression estimator for time series data. It is an area that has been investigated fairly thoroughly by many authors. A partial list of related works includes Robinson (1983), Roussas (1990), Troung and Stone (1992), Tran (1993) and Liebscher (1999). While these results are well established theoretically, one can only say from these results that some extensions of the *iid* result to weak dependence is possible. In this short article theoretical results such as consistency and rates of convergence are derived, which addresses some insights into the problem such as dependence effect on the estimator. Consider a stationary sequence of pairs of random variables $\{(X_j, Y_j) \mid j \geq 1\}$, where the X's are $\mathbb{R}^d$-valued and the Y's are real-valued. The problem under consideration is providing a nonparametric estimator of the conditional expectation $\mathbb{E}(Y|X = x)$. In this problem one of the most popular nonparametric methods is the kernel regression estimator given by

$$r_n(x) = R_n(x)/f_n(x) \tag{1.1}$$

where

$$R_n(x) = \left(nh^d\right)^{-1} \sum_{i=1}^{n} Y_i \, K\left(\frac{x - X_i}{h}\right)$$

Received October 2001; accepted May 2002.
†This work was supported by Korea Research Foundation Grant (KRF 2000-015-DP0052).
[1]Department of Statistics, Keimyung University, Taegu 704-701, Korea
[2]R&D Department, LKFS, Seoul 135-010, Korea

and

$$f_n(x) = \left(nh^d\right)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

and $K > 0$ is the kernel function (see also the assumptions A1 and A2 below). In our theoretical work the uniform strong consistency of r_n and its specific rate are studied. From this we will remark some dependence effect on the convergence rate of the estimator and its smoothing parameter h. Behind this is that dependency affects the performance of estimator by slowing down the speed of convergence.

The following point, however, should be pointed out. Though most of papers handling this setting include the well-known AR (auto-regression) model by taking $(X_i, Y_i) = (X_i, X_{i+1})$, it is our intention not to include the AR model here because AR appears to require a different treatment. See Remark 2.2 below the main theorems. This paper consists of 3 sections. Section 2 contains theoretical results and discussions, and the proofs will be given in Section 3.

2. Theoretical Results

We start this section with the following well-known definition of α-mixing. Consider the σ algebras given by

$$\mathcal{M}_m^n = \sigma\Big(\xi(t) : m \le t < n\Big), \quad -\infty \le m < n \le \infty,$$

with a clear change for $m = -\infty$. The α-mixing coefficients are defined by

$$\alpha(t) = \sup\Big\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{M}_{-\infty}^0 \text{ and } B \in \mathcal{M}_t^\infty\Big\}.$$

A process is α- mixing if $\alpha(t) \to 0$ as $t \to \infty$, but it is generally advisable to impose some rate on this convergence of the strong mixing coefficients to zero to reach useful results. See Roussas (1990) for the advantages of assuming α-mixing condition.

Set

$$r(x) = \mathbb{E}\,(Y|X = x), \quad x \in \mathbb{R}^d.$$

Notice that the quantity $r(x)$ is to be estimated by $r_n(x)$. Now the following assumptions are listed :

(A1) $\int K(x)\,dx = 1$ and for some $\alpha > 0$,

$$\int_{\mathbb{R}^d} \| u \|^\alpha K(u)\,du < \infty.$$

(A2) K is Hölder continuous with the exponent $\alpha > 0$ given in (A1), *i.e.*, there is a constant C such that for all $x, y \in \mathbb{R}^d$,

$$|K(x) - K(y)| \le C \parallel x - y \parallel^\alpha .$$

(A3) f is Hölder continuous with the same exponent $\alpha > 0$ as in (A2) and is bounded away from zero on a compact set $S \subseteq \mathbb{R}^d$ (*i.e.*, $\inf f(x) \ge \delta > 0$).

(A4) $\psi(x) = r(x)f(x)$ is Hölder continuous with the same exponent $\alpha > 0$ as in (A3).

(A5) For all $s > 0$,

$$\mathbb{E}\, |Y|^s < \infty.$$

(A6) The bandwidth $h > 0$ is a sequence satisfying $h \sim cn^{-\theta}$ for some positive constants c and $\theta < 1/d$.

(A7) Let $\mu = \theta(d + \alpha)/\alpha$. The mixing coefficients $\alpha(i)$ satisfy

$$\alpha(i) \le Ci^{-p}$$

for some positive constants C and p where p satisfies

$$p > \frac{9(1 + d\mu)\theta d}{2(1 - \theta d)^2}. \tag{2.1}$$

Then the following strong consistency result is established first.

Theorem 2.1. *Under assumptions (A1) through (A7),*

$$\sup_{x \in S} |r_n(x) - r(x)| \to 0, \ a.s.$$

Now convergence rates are investigated under dependency. First, the local pointwise rate of convergence is provided in the following result.

Theorem 2.2. *Assume that*

$$\alpha(i) \le Ci^{-p} \tag{2.2}$$

for some positive constants C and $p > 1$. Suppose that $h \sim n^{-1/[d(1+1/p)+2]}$ and the assumptions (A1) - (A5) hold with $\alpha = 1$. Then

$$|r_n(x) - r(x)| = O_p\left(n^{-1/[d(1+1/p)+2]}\right), \quad x \in S. \tag{2.3}$$

The L_∞ rate of convergence is supplied in the following result.

Theorem 2.3. *Suppose that $h \sim n^{-1/(2+d)}$ and assumptions (A1) - (A5) hold with $\alpha = 1$. Further assume that*

$$\alpha(i) \leq C\rho^i$$

for some positive constants C and $\rho \in (0,1)$. Then for any δ such that $0 < \delta < 1/(2+d)$,

$$\sup_{x \in S} |r_n(x) - r(x)| = o\left(n^{-\delta}\right) \quad a.s.$$

Remark 2.1. Our theorems notice an implicit underlying relation between the amount of dependence and the size of bandwidth h. First notice that a small θ verifying $\theta d < 1$ in Theorem 2.1 results in a small p while a large θ produces a large p. Recall that p is the exponent of the mixing coefficients. For example, when $\alpha = 1$, $d = 1$ and $\theta = 0.5 - \epsilon$ for some small positive ϵ our assumption (A7) yields $p > 18$, but taking θ close to 0, which is a weaker condition, a small $p > 1$ is achieved. Theorem 2.2 suggests a similar result in terms of convergence rate. Indeed their optimal bounds for the convergence rate are shown to slow down by the amount of dependence. In addition the optimal bandwidth $h(\to 0)$ found from the bounds seems to be larger than the *iid* optimal bandwidth h.

Theorem 2.3 implies that the almost identical result to the *iid* case holds for geometric mixing case. Note that the *iid* pointwise optimal rate is $O\left(n^{-2/(d+4)}\right)$ and the *iid* L^∞ optimal rate is $O\left((\log n/n)^{2/(d+4)}\right)$. In addition it should be mentioned that similar results to Theorem 2.3 have been established under geometric mixing with various technical conditions (see, *e.g.*, Theorems 3.2 and 3.3 of Bosq, 1996).

Remark 2.2. As noted in introduction, we believe the above remark does not apply to the well-known AR model. The reason is presented simply as follows. When X_t and X_{t+1} are independent, the auto-regression function is a constant (*i.e.*, $\mathbb{E}(X_{t+1}|X_t = x) = \mathbb{E}(X_{t+1}) = \mu$) which could be best estimated by choosing $h = \infty$ in r_n. This contrasts with our findings from theorem that a relatively large optimal bandwidth h is desirable to handle the dependence in recovering $\mathbb{E}(Y|X = x)$. Thus in order to investigate dependence effect on bandwidth h for AR model, a different approach seems to be necessary and further research is certainly desirable in this direction.

TABLE 2.1 *Simulation results*

	$\phi = 0$		$\phi = 0.10$		$\phi = 0.95$	
	h	ASE	h	ASE	h	ASE
Mean	0.2198	6.134e-07	0.24160	7.532e-07	0.3315	2.077e-05
SD	0.0565	3.242e-07	0.09211	5.241e-07	0.1669	4.366e-05
Min	0.1257	1.764e-07	0.09934	2.822e-07	0.1556	2.558e-07
Max	0.3191	1.311e-06	0.38220	1.564e-06	0.6230	1.443e-04

Remark 2.3. In order to check the theoretical results of theorems, a small simulation work is done. Indeed (X_i, Y_i), $i = 1, \cdots, 20$ is generated by $X_i \sim$ Uniform$(0, 1)$ and $Y_i = m(X_i) + \epsilon_i$ where $m(x) = x^3(1-x)^3$ and $\epsilon_i = \phi\epsilon_{i-1} + \gamma_i$ with the *iid* $\gamma_i \sim N(0, 0.0015^2)$. The result summarized in Table 2.1 shows that as ϕ approaches from 0 to 0.95 (*i.e.*, as dependence gets more severe), the optimal bandwidth h and ASE tend to increase, which is an explicit suggestion from Theorems 2.1 and 2.2. In the simulation study we generate a geometric strong mixing process instead of the algebraic α-mixing process with a varying exponent p assumed in Theorems 2.1 and 2.2. The reason is that there is not available technique for generating such an α-mixing process as far as we know. Note that a relatively small sample size $n = 20$ is used because with increase of n the dependence effect is to be washed out as Theorem 2.3 suggested.

3. Proofs

The following is our main mathematical tool due to Cox and Kim (1995).

Lemma 3.1. *Let $\xi(t)$ be a strong mixing process and let l be a positive integer. Assume that $\mathbb{E}\,\xi(t) = 0$, and that for some $q > 2$*

$$M_{ql} = \sup_t \left\{ \parallel \xi(t) \parallel_{ql} \right\} = \sup_t \left\{ \left(\mathbb{E}\ |\xi(t)|^{ql} \right)^{1/(ql)} \right\} \leq 1.$$

Suppose further that there is a constant ν not depending on t such that

$$\mathbb{E}\ |\xi(t)|^k \leq \nu, \quad 2 \leq k \leq 2l.$$

Finally, assume that the mixing coefficients satisfy

$$\sum_{i=1}^{\infty} i^{l-1} \alpha(i)^{1-2/q} < \infty.$$

Then there exists a constant C depending on l but not depending on the distribution of $\xi(t)$ nor on ν, n, nor on P such that

$$\mathbb{E}\left(\sum_{i=1}^{n}\xi(i)\right)^{2l} \le C\left\{n^l M_{ql}^{2l}\sum_{i=P}^{\infty} i^{l-1}\alpha(i)^{1-2/q} + \sum_{j=1}^{l} n^j P^{2l-j}\nu^j\right\}$$

for any integers n and P with $0 < P < n$.

Proof of Theorem 2.1. Set

$$R_n(x) = \left(nh^d\right)^{-1}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h}\right).$$

Observe that

$$\begin{aligned}|r_n(x) - r(x)| &= f_n(x)^{-1}|R_n(x) - r(x)f_n(x)| \\ &\le f_n(x)^{-1}\Big(|R_n(x) - \mathbb{E}_n(x)| + |\mathbb{E}_n(x) - \psi(x)| \\ &\qquad + r(x)|f_n(x) - f(x)|\Big) \qquad (3.1)\end{aligned}$$

where it is recalled that $\psi(x) = r(x)f(x)$. The behavior of the terms on the right hand side of equation (3.1) is exhibited in the below. Under the conditions of theorem, Kim and Cox (1996) showed that

$$\sup_{x\in S}|f_n(x) - f(x)| \to 0, \ a.s. \qquad (3.2)$$

Since $\sup_{x\in S} f_n(x)^{-1}$ is bounded *a.s.* for all sufficiently large n by assumption (A3) and equation (3.2), it suffices to show that the terms inside the brackets of equation (3.1) converge to zero uniformly *a.s.*

Since r is continuous, one obtains from equation (3.2) that under the conditions of theorem

$$\sup_{x\in S} r(x)|f_n(x) - f(x)| \to 0, \ a.s. \qquad (3.3)$$

Next we assert that $\mathbb{E}\,R_n(x) - \psi(x)$ tends to zero uniformly in $x \in S$. By assumptions (A1), (A3) and (A4)

$$\begin{aligned}|\,\mathbb{E}\,R_n(x) - \psi(x)| &= \left|\int K(u)\Big(\psi(x-hu) - \psi(x)\Big)du\right| \\ &\le \int |K(u)|\,\|\,u\,\|^{\alpha}\,du\ \ h^{\alpha} \\ &\le Ch^{\alpha}, \qquad (3.4)\end{aligned}$$

where C denotes a generic positive constant. Since the latter bound tends to zero uniformly over $x \in S$ as $n \to \infty$ by assumption (A6), this proves the assertion.

Now it will be essential to show that $R_n - \mathbb{E}_n$ converges to zero uniformly *a.s.* Working as in the proof of Theorem 1 of Kim and Cox (1996), we consider sequences of finite subsets $S'_n \subseteq S$ whose cardinality grows algebraically fast and prove the uniform convergence on the subsets. The rest of S will be treated by a Hölder continuity argument. Let $\eta > 0$ be given for now and S'_n be such that for all $x \in S$, there exists $x' = x'(x) \in S'_n$ such that $\| x' - x \| < n^{-\eta}$. Then one may assume

$$\#S'_n \le Cn^{d\eta},$$

where $\#$ denotes the cardinality. For a given $\epsilon > 0$, set

$$p_n = P\left(\sup_{x' \in S'_n} |R_n - \mathbb{E}_n| > \epsilon\right)$$

and put

$$\xi(i) = Y_i\, K\Big((x' - X_i)/h\Big) - \mathbb{E}\, Y_i\, K\Big((x' - X_i)/h\Big).$$

Then $\mathbb{E}\,\xi(i) = 0$. Let l be a positive integer to be determined. Using Lemma 3.1 with $\nu = Ch^d$ and $q = \infty$ (see assumption (A5)) and following the argument of the proof of Theorem 2.2 below (see equations (3.10)–(3.12) below), one has

$$\begin{aligned} p_n &\le Cn^{d\eta}\epsilon^{-2l}(nh^d)^{-2l}\left(n^l \sum_{i=P}^{\infty} i^{l-1}\alpha(i) + \sum_{j=1}^{l} n^j P^{2l-j} h^{dj}\right) \\ &\le Cn^{d\eta + l(-1+\theta d + \theta dl/p)}. \end{aligned}$$

Now our goal is to choose l so that $\sum p_n < \infty$. For this it suffices to choose a positive integer l such that

$$l\Big(1 - \theta d(1 + l/p)\Big) > d\eta + 1. \tag{3.5}$$

By solving the above quadratic inequality in l, one can show that equation (3.5) holds provided

$$p > \frac{9(1 + d\eta)\theta d}{2(1 - \theta d)^2}. \tag{3.6}$$

Extension of the result to the entire interval S is made by a Hölder continuity argument. We have

$$\sup_{x \in S} |R_n(x) - \psi(x) - R_n(x') + \psi(x')| \le \sup |R_n(x) - R_n(x')| + Cn^{-\eta\alpha},$$

where we have used assumption (A4) at the last step. In the above, $x' = x'(x)$ is the point in S_n' closest to x. There remains only to show that the first term in the last expression converges to zero *a.s.* By Hölder continuity of K, one may easily see that under the conditions of Theorem 2.1

$$\sup |R_n(x) - R(x')| \leq C n^{\theta d + (-\eta+\theta)\alpha}. \tag{3.7}$$

Then the bound in equation (3.7) converges to zero provided

$$\eta > \mu. \tag{3.8}$$

This completes the extension from S_n' to S.

To tie it all up, because of assumption (A7) we can choose η satisfying both equations (3.6) and (3.8) using a continuity argument. This finishes the proof. □

Proof of Theorem 2.2. Using equation (3.4) with $\alpha = 1$, one has

$$|\,\mathbb{E}\, R_n(x) - \psi(x)| \leq Ch. \tag{3.9}$$

And an application of Lemma 3.1 yields

$$p_{n1} = P\Big(|R_n(x) - \mathbb{E}\, R_n(x)| > \epsilon\Big) \leq C n^{-1} h^{-d(1+1/p)}. \tag{3.10}$$

Indeed by Chevyshev inequality and Lemma 3.1 with $l = 1$, $q = \infty$ and $\nu = Ch^d$, p_{n1} is bounded above by

$$C\left(nh^d\right)^{-2}\left(nP^{1-p} + nPh^d\right) = C\left(n^{-1}h^{-2d}P^{1-p} + n^{-1}Ph^{-d}\right), \tag{3.11}$$

provided $0 < P < n$ and $p > 1$. Then the best pick for P will be made by solving $n^{-1}h^{-2d}P^{1-p} = n^{-1}Ph^{-d}$, which gives $P = h^{-d/p}$. When this is put back into equation (3.11), equation (3.10) follows. Thus

$$|R_n(x) - \mathbb{E}_n(x)| = O_p\left(n^{-1/2}h^{-d(1+1/p)/2}\right). \tag{3.12}$$

Note that Theorem 2 of Kim and Cox (1996) proved

$$|f_n(x) - f(x)| = O_p\left(n^{-1/[d(1+1/p)+2]}\right). \tag{3.13}$$

Now by plugging back equations (3.9), (3.12) and (3.13) in equation (3.1), one has

$$|r_n(x) - r(x)| = O_p\left(n^{-1/2}h^{-d(1+r/p)/2} + h\right).$$

By choosing $h \sim n^{-1/[d(1+r/p)+2]}$, the desired result follows. □

Proof of Theorem 2.3. Under the conditions of theorem, Kim and Cox (1996) showed that

$$\sup_{x \in S} |f_n(x) - f(x)| = o\left(n^{-\delta}\right) \quad a.s., \tag{3.14}$$

for any $0 < \delta < 1/(2+d)$. Then as working in the proof of Theorem 2.1,

$$\sum_n P\left(n^\delta \sup_{x \in S} |r_n(x) - r(x)| > \epsilon\right) < \infty$$

provided

$$\delta < \theta \tag{3.15}$$

and

$$2l\delta + \theta d(d+1) + l\theta d(1 + l/p) < l - 1. \tag{3.16}$$

Obviously the condition (3.15) comes from equation (3.4) and the condition (3.16) comes from equations (3.5) and (3.8). Therefore, assuming algebraic mixing process (*i.e.*, $\alpha(i) = O(i^{-p})$) for now,

$$\delta + \frac{\theta d(1 + l/p)}{2} + \frac{\theta d(d+1)}{2l} < \frac{l-1}{2l}, \tag{3.17}$$

for some positive integer $l > 1$. Since we are dealing with geometric mixing process, we may take $p = \infty$ and by letting $l \to \infty$, equation (3.17) becomes

$$\delta + \frac{\theta d}{2} < \frac{1}{2}.$$

Hence for the desired convergence rate one must have

$$0 < \delta < \frac{1}{2} - \frac{\theta d}{2}. \tag{3.18}$$

From equations (3.15) and (3.18), one may choose $\theta = 1/(2+d)$ by solving $\theta = 1/2 - \theta d/2$. Finally one obtains

$$\sup_{x \in S} |r_n(x) - r(x)| = o\left(n^{-\delta}\right) \quad a.s.,$$

for any $\delta < 1/(2+d)$. □

Acknowledgements

The authors are very grateful to the Associate Editor and anonymous referees for very helpful comments which led to the present improved version of the article.

References

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes*, Springer-Verlag, New York.

Cox, D. and Kim, T. Y. (1995). "Moment bounds for mixing random variables useful in nonparametric function estimation", *Stochastic Processes and Their Applications*, **56**, 151–158.

Kim, T. Y. and Cox, D.(1996). "Uniform strong consistency of kernel density estimators under dependence", *Statistics & Probability Letters*, **26**, 179–185.

Robinson, P. M. (1983). "Nonparametric estimators for time series", *Journal of Time Series Analysis*, **4**, 185–207.

Roussas, G. (1990). "Nonparametric regression estimation under mixing conditions", *Stochastic Processes and Their Applications*, **36**, 107–116.

Tran L. H. (1993). "Nonparametric function estimation for time series by local average estimators", *The Annals of Statistics*, **21**, 1040–1057.

Troung, Y. K. and Stone, C. J. (1992). "Nonparametric function estimation involving time series", *The Annals of Statistics*, **20**, 77–97.