Commun. Korean Math. Soc. **19** (2004), No. 2, pp. 345–354

ROBUST REGRESSION SMOOTHING FOR DEPENDENT OBSERVATIONS

TAE YOON KIM, GYU MOON SONG AND JANG HAN KIM

ABSTRACT. Boente and Fraiman [2] studied robust nonparametric estimators for regression or autoregression problems when the observations exhibit serial dependence. They established strong consistency of two families of M-type robust equivariant estimators for ϕ-mixing processes. In this paper we extend their results to weaker α-mixing processes.

1. Introduction

For regression problem with serially dependent observations, various nonparametric techniques have been used for recovering the unknown regression function. Two most popular methods among them are kernel and k-nearest methods, presented by Nadaraya [7] and Watson [10] and by Collomb [3] respectively. Early results for this problem include Collomb [3], Robinson [8] and Doukhan, Leon and Portal [6] as they revealed asymptotic properties of the related estimators and predictors. Evidently both of kernel methods and k-nearest kernel methods are weighted averages of the response variables and therefore are highly sensitive to large fluctuations in the data. Thus robust estimators obtained via M-estimates have been considered by several authors. See Robinson [9], Collomb and Härdle [4] and Boente and Fraiman [2]. Indeed Robinson [9] adapted the robust M-estimators of a location parameter with kernel weights to time series model and established a central limit theorem for such estimators when scale is known. A similar approach was taken by Collomb and Härdle [4] who established uniform convergence of this family of estimators for ϕ-mixing processes.

Received April 9, 2003.
2000 Mathematics Subject Classification: 62M10, 62G07.
Key words and phrases: Robust nonparametric regression, strong consistency, α-mixing sequence.
This research is supported by KRF Grant 2003-070-C00008.

Later Boente and Fraiman [2] consider robust scale equivariant nonparametric M-estimators based on kernel methods and k-nearest neighbor kernel methods for which they obtain strong pointwise convergence for ϕ-mixing processes. In this paper we follow the approach developed by Boente and Fraiman [2] and extend their results to the weaker α-mixing case. Indeed their strong consistency and asymptotically strong robustness (ASR) is established when the underlying sequence exhibits α-mixing.

Let $\{(X_t, Y_t) : t \geq p+1\}$ be a strictly stationary process, $X_t \in R^p$ and $Y_t \in R$. For $x \in R^p$ let $\phi(x) = E(Y_t|X_t = x)$. The Nadaraya-Watson regression estimator is given by

$$\phi_T(x) = \sum_{t=p+1}^{T} w_{tT}(x)Y_t,$$

where

$$(1.1)\qquad \begin{aligned} w_{tT}(x) &= w_{tT}(x, X_{p+1}, \ldots, X_T) \\ &= K((X_t - x)/h_T) / \sum_{\tau=p+1}^{T} K((X_\tau - x)/h_T), \end{aligned}$$

with K a nonnegative integrable function on R^p and $h_T > 0$ which can be also used for a p-th order autoregressive model, i.e., a strictly stationary real valued process $\{Z_t : t \in N\}$ satisfying

$$(1.2)\qquad Z_t = g(X_t) + e_t$$

where $X_t = (Z_{t-1}, \ldots, Z_{t-p})$, $Y_t = Z_t$, e_t is independent of $\{Z_{t-1}, Z_{t-2}, \ldots\}$ and $E(e_t) = 0$.

To detail our robust problem let (X, Y) be a random vector with the same distribution as (X_t, Y_t). Then the robust conditional location functional $g(X) = E^{\psi}(Y|X)$ defined in Boente and Fraiman [1] is the essentially unique $\sigma(X)$-measurable function $g(X)$ that verifies

$$(1.3)\qquad E\{h(X)\psi[(Y - g(X))/s(X)]\} = 0$$

for all integrable function h, where $\sigma(X)$ is the σ-algebra generated by X, $s(X)$ is a robust measure of the conditional scale, e.g.,

$$(1.4)\qquad s(x) = med(|Y - m(x)|) = MAD_c(x),$$

$m(x) = med(Y|X = x)$ is the median of a regular version $F(y|X = x)$ of the conditional distribution function and $\psi : R \to R$ is a strictly increasing, bounded and continuous function. When the distribution of $Y|X = x$ has half or more than half of its mass at one single point we redefine $s(x) = 1$. If the conditional distribution function $F(y|X = x)$ is symmetric around $\phi(x)$ and ψ is odd, we have $g(x) = \phi(x)$. Then, in this sense, it is a natural extension of the conditional expectation $E(Y|X)$.

In Theorem 2.1 of Boente and Fraiman [1], it was shown that the solution of (1.3) exists, is unique and measurable. The weak continuity of the functional so defined was proved in Theorem 2.2 there. Then we obtain consistent and asymptotically strongly robust (ASR) estimates of the autoregression function by applying the functional so defined to estimates $F_T(y|X = x)$ of $F(y|X = x)$, verifying that $F_T(y|X = x) \to_\omega F(y|X = x)$ as $T \to \infty$ a.s. (μ), where $\to_\omega$ stands for weak convergence and μ denotes the marginal distribution of the vector X. See Boente and Fraiman [1] for detailed definition of ASR. Now we will consider the two families of estimators of $F(y|X = x)$ considered by Boente and Fraiman [2].

1. Estimators based on kernel weights. These are defined by

$$F_T(y|X = x) = \sum_{t=p+1}^{T} w_{tT}(x) I_A(Y_t), \tag{1.5}$$

where $A = (-\infty, y]$, and I_A denotes the indicator function of the set A and w_{tT} is defined in (1.1).

2. Estimators based on k-nearest neighbor kernel methods. These are defined by

$$\hat{F}_T(y|X = x) = \sum_{t=p+1}^{T} \hat{w}_{tT}(x) I_A(Y_t), \tag{1.6}$$

where

$$\begin{aligned} \hat{w}_{tT}(x) &= \hat{w}_{tT}(x, X_{p+1}, \ldots, X_T) \\ &= K((X_t - x)/H_T) / \sum_{\tau=p+1}^{T} K((X_\tau - x)/H_T), \end{aligned}$$

H_T is the distance between x and the k-nearest of x among $X_{p+1}, \ldots, X_T$ and $k = k_T$ is a fixed integer. In particular, when $K(t) = I_{\|u\|<1}(t)$, where $\|\cdot\|$ is any norm on R^p, we obtain the uniform $k-NN$ estimate.

Denote by $s_T(x)$ and $\hat{s}_T(x)$ the scale measures corresponding to $F_T(y|X = x)$ and $\hat{F}_T(y|X = x)$, respectively as defined in (1.4). The corresponding robust nonparametric estimates of $g(x)$ are given by the unique solution of

$$\sum_{t=p+1}^{T} w_{tT}\ \psi[(Y_t - g_T(x))/s_T(x)] = 0 \tag{1.7}$$

and

$$\sum_{t=p+1}^{T} \hat{w}_{tT}\ \psi[(Y_t - \hat{g}_T(x))/\hat{s}_T(x)] = 0. \tag{1.8}$$

2. Main results

Before stating the main results, the following set of assumptions are listed.

A1. $\psi : R \to R$ is a strictly increasing, bounded and continuous function such that $\lim_{u\to\infty} \psi(u) = a > 0$ and $\lim_{u\to-\infty} \psi(u) = b < 0$.

A2. Either of the following statements holds.

(a) $s(x)$ is given by a functional which is weakly continuous at F, for almost all x.

(b) ψ is odd and $F(y|X = x)$ is symmetric around $g(x)$ and a continuous function of y for each fixed x.

H1. The process $\{(X_t, Y_t) : t \geq p+1\}$ is an α-mixing, i.e. there exists a non-increasing sequence of numbers $\{\alpha(n) : n \in N\}$ with $\lim_{n\to\infty} \alpha(n) = 0$ such that for any integer n,

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(n),$$

where $A \in M_{p+1}^{t}$, $B \in M_{t+n}^{\infty}$ and M_u^v is the σ-field generated by the random vectors $\{(X_t, Y_t) : u \leq t \leq v\}$. We also assume: the mixing coefficients decay algebraically fast, i.e., there exist $w > 1$ and $a > 0$ such that $\alpha(n) \leq an^{-w}$.

H2. $K : R^p \to R$ is a bounded nonnegative function satisfying

$$aI_{\|u\|\leq r}(u) \leq K(u) \text{ for some } a > 0,\ \ r > 0,$$

$$a_1 H(\| u \|) \leq K(u) \leq a_2 H(\| u \|),$$

where a_1 and a_2 are positive numbers and $H : R^+ \to R^+$ is bounded, decreasing and such that $t^p H(t) \to 0$ as $t \to \infty$.

H3. The sequences $\{h_T : T \in N\}$ is such that

$$h_T \to 0 \text{ and } Th_T^p \to \infty \text{ as } T \to \infty.$$

H4. Let $h_T \sim T^{-\theta}$ for some $\theta > 0$. Then there exists a positive integer $l < w$ such that

$$0 < \theta < (1 - 1/l)/[p(1 + l/w)]. \tag{2.1}$$

THEOREM 2.1. *Assume A1, A2 and H1-H4 hold. Then we have*
(a) $g_T(x) \to g(x)$ *a.s. as* $T \to \infty$ *for almost all* $x(\mu)$.
(b) $g_T(x)$ *is asymptotically strongly robust (ASR) at* μ.

REMARK 2.1. Recall that Theorem 2.2 of Boente and Fraiman [1] entails that Theorem 2.1 above is a consequence of the almost everywhere weak convergence of $F_T(y|X = x)$ to $F(y|X = x)$ established in Theorem 3.1 below. Boente and Fraiman [2] have obtained their consistency results for geometric α-mixing in which they require h_T to satisfy $T^{1/4}(h_T^p)^{(1+\delta)/4}/\log T \to \infty$ for some $\delta > 0$ as $T \to \infty$. Then assuming $h_T \sim T^{-\theta}$, their result reduces to $\theta < 1/[p(1 + \delta)]$. In the meantime since we may take $w = \infty$ and then let $l \to \infty$ in (2.1) for the geometric α-mixing, our result could give a weaker condition on the bandwidth selection, i.e., $0 < \theta < 1/p$.

REMARK 2.2. One may give a sufficient condition for (2.1) to hold. Indeed one may show that if

$$w > \frac{9\theta p}{2(1 - \theta p)^2} \tag{2.2}$$

then (2.1) hold. This reveals an interesting relationship between the dependence structure and the rates of convergence of h_T. In fact, a small θ satisfying $\theta p < 1$ produces a small w while a large θ produces a large w. Remember w is the exponent of the mixing coefficients. In other words, we need to let $h \to 0$ more slowly as the dependence becomes severe. This makes sense because for the severely dependent data, each data point itself may not be reliable but a group of data close in distance may be more reliable.

As noted by Boente and Fraiman [2], our results do not require any restriction on the probability distribution μ of the vector X. Hence the result obtained are robust and distribution-free in the sense that they are true for all μ.

For the case of the k-nearest neighbor kernel methods, we will replace H2, H3 and H4 by the following assumptions.

B1. The vector X has a density $f(x)$. $K : R^p \to R$ is a bounded nonnegative function, $\int K(u)du = 1$ and $K(u) \leq c_1 I_{\{\|u\| \leq r\}(u)}(u)$.

B2. The sequence $\{k_T : t \in N\}$ satisfies $k_T \to \infty$ and $k_T/T \to 0$ as $T \to \infty$.

B3. Let $k_T \sim T^\gamma$ for some $0 < \gamma < 1$. Then there exists a positive integer $l < w$ such that

$$(1/l + l/w)/(1 + l/w) < \gamma. \tag{2.3}$$

B4. $K(uz) \geq K(z)$ for all $u \in (0,1)$.

THEOREM 2.2. *Under A1, A2, H1 and B1-B4 we have that:*
(a) $\hat{g}_T(x) \to g(x)$ *a.s. as* $T \to \infty$ *for almost all* $x(\mu)$.
(b) $\hat{g}_T(x)$ *is ASR at* μ.

REMARK 2.3. As for the kernel weights estimators, the above result follows from the almost everywhere weak convergence of $\hat{F}_T(y|X = x)$ to $F(y|X = x)$ established in Theorem 3.2 below. Since $\theta p = 1 - \gamma$, discussions in the remarks 2.1-2.2 may be adjusted in obvious way. For example, one may deduce (2.3) from (2.1) using $\theta p = 1 - \gamma$. In a similar fashion, one may easily notice that for the same geometric α-mixing, taking $w = \infty$ and then letting $l \to \infty$ reduces (2.3) to $\gamma > 0$, while Boente and Fraiman [2] leads to $\gamma > \delta$ for some $\delta > 0$. Also for (2.3) to hold one may provide a sufficient condition

$$w > \frac{9(1-\gamma)}{2\gamma^2}.$$

Since a small $\gamma < 1$ yields a large w above, similar observations can be made about the convergence rate on k_T in the k-nearest kernel methods.

3. Estimating the conditional distribution function

In this section we will study the strong consistency of $F_T(y|X = x)$ and $\hat{F}_T(y|X = x)$, defined by (1.5) and (1.6) respectively.

Given a Borel set $A \subset R$ we denote by $\phi_T(x)$ and $\phi(x)$ the function

$$\phi_T(x) = \sum_{t=p+1}^{T} w_{tT}(x) I_A(Y_t),$$

$$\phi(x) = E(I_A(Y)|X = x),$$

where w_{tT} is defined in (1.1).

THEOREM 3.1. *Assume H1-H4. Then:*
(i) $\phi_T(x) \to \phi(x)$ *a.s. for almost all* x.
(ii) $\lim_{T\to\infty} \sup_y |F_T(y|X = x) - F(y|X = x)| = 0$ *a.s. for almost all* x.

PROOF. (i) According to Boente and Fraiman [2], the verification of (i) will be complete if we show that

$$\frac{1}{T}\sum_{t=p+1}^{T}(\eta_{tT} - E(\eta_{tT})) = \frac{S_T}{T} \to 0 \text{ a.s. as } T \to \infty \tag{3.1}$$

where

$$\eta_{tT} = K(\frac{X_t - x}{h})/a_T \text{ and } a_T = EK(\frac{X - x}{h}).$$

For the verification of (3.1), we will use the following moment bounds due to Cox and Kim (1994).

LEMMA 3.1. *Let $\xi(t)$ be a strong mixing process. Let l be a positive integer and assume $E\xi(t) = 0$, and that for some $q > 2$*

$$M_{qr} = \sup_t\{\| \xi(t) \|_{ql}\} = \sup_t\{(E|\xi(t)|^{ql})^{1/(ql)}\} \leq 1. \tag{3.2}$$

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

$$E[|\xi(t)|^k] \leq \nu, \quad 2 \leq k \leq 2l.$$

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 P such that

$$E[(\sum_{i=1}^{n}\xi(i))^{2l}] \leq C\{n^r M_{qr}^{2l}\sum_{i=P}^{\infty} i^{l-1}\alpha(i)^{1-2/q} + \sum_{j=1}^{l} n^j P^{2l-j}\nu^j\}$$

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

Let $\xi(t) = I_A(Y_t)K((X_t - x)/h) - EI_A(Y_t)K((X_t - x)/h)$. Then $E\xi(t) = 0$. Let l be a positive integer to be determined. Then if $2 \leq k \leq 2l$,

$$E|\xi(t)|^k \leq Ca_T.$$

Similarly, one may show that for any $q > 2$,

$$\| \xi(t) \|_{ql} \leq Ca_T^{1/(ql)}.$$

Thus, (3.2) will hold for T sufficiently large. Thus, we may apply Lemma 3.1 with $\nu = Ca_T$, and then

$$E(\sum_{i=1}^{T}\xi(i))^{2l} \leq C\left[T^r a_T^{2/q}\sum_{i=P}^{\infty} i^{l-1}\alpha(i)^{1-2/q} + \sum_{j=1}^{l} T^j P^{2l-j} a_T^j\right].$$

For convenience, let $s = 2/q$ so that $0 < s < 1$. By Markov's inequality we obtain

$$p_T = P\left(\left|\frac{S_T}{T}\right| > \epsilon\right) \le (Ta_T)^{-2r} E[\sum_{t=p+1}^{T} \xi(t)]^{2l}$$

$$\le C(Ta_T)^{-2l}\left[T^r a_T^s \sum_{i=P}^{\infty} i^{l-1}\alpha(i)^{1-s} + \sum_{j=1}^{l} T^j P^{2l-j} a_T^j\right]. \tag{3.3}$$

Now by (3.2),

$$\sum_{i=P}^{\infty} i^{l-1}\alpha(i)^{1-s} \le CP^{l-w(1-s)},$$

provided

$$0 < P < T, \tag{3.4}$$

and

$$w(1-s) > l. \tag{3.5}$$

Substituting this back into (3.3) yields

$$p_T \le C(Ta_T)^{-2r}\left[T^r a_T^s P^{l-w(1-s)} + \sum_{j=1}^{l} T^j P^{2l-j} a_T^j\right]$$

$$= C[T^{-l} a_T^{-2l+s} P^{l-w(1-s)} + \sum_{j=1}^{l} (T^{-1} P a_T^{-1})^{2l-j}]. \tag{3.6}$$

In the above one may easily see that the best choice for P will be obtained by solving

$$a_T^{-2l+s} P^{l-w(1-s)} = a_T^{-l} P^r,$$

which gives

$$P = a_T^{-(l-s)/[w(1-s)]}. \tag{3.7}$$

Clearly (3.4) holds by (3.5). When this is put back into (3.6) we obtain

$$p_T \le CT^{-l} a_T^{-l(1+(l-s)/[w(1-s)])} \le CT^{-r}\mu(S_{uh})^{-l(1+(l-s)/[w(1-s)])}$$

$$\le CT^{-l}(h^p)^{-l(1+(l-s)/[w(1-s)])} \le CT^{-l}(T^{-\theta p})^{-l(1+(l-s)/[w(1-s)])} \tag{3.8}$$

where S_{uh} is the closed ball of radius uh centered at x. The last result follows from (H4) as $h^p u^p/\mu(S_{uh}) \to (d\lambda_1/d\mu)(x)$, where λ_1 is the μ absolutely continuous part of the Lebesgue measure on R^p. Now our goal is to choose r and s so that $\sum_T p_T < \infty$. By a continuity argument it will be sufficient to choose a positive integer l such that

$$r[1 - \theta p(1 + l/w)] > 1 \tag{3.9}$$

for then we make choose $s > 0$ sufficiently close to 0 to obtain (3.8). Then (3.9) holds by H4 and (3.1) follows. Proof of (i) is complete.
(ii) Proof follows from (i) by an argument similar to the one used to prove the Glivenko-Cantelli Theorem. □

THEOREM 3.2. *Under H1 and B1-B4 we have that:*
(i) $\hat{\phi}_T(x) \to \phi(x)$ *a.s. for almost all* x.
(ii) $\lim_{T\to\infty} \sup_y |\hat{F}_T(y|X = x) - F(y|X = x)| = 0$ *a.s. for almost all* x.

PROOF. We will just sketch the proof here because proof can be done by proceeding exactly as in section 4.2 of Boente and Fraiman [2]. Indeed Lemmas 4.4 and 4.7 of Boente and Fraiman continue to hold under the conditions of Theorem. And their Lemmas 4.5 and 4.6 may be established by using our Lemma 3.1 above as in the proof of Theorem 3.1 which completes the proof. □

References

[1] G. Boente and R. Fraiman, *Robust nonparametric regression estimation*, J. Multivariate Anal. **29** (1989), 180–198.

[2] G. Boente and R. Fraiman, *Robust nonparametric regression estimation for dependent observations*, Ann. Statist. **17** (1989), 1242–1256.

[3] G. Collomb, *Propriétés de convergence presque complète du prédicteur à noyau*, Zeit. Wahr. ver. Geb. **66** (1984), 441–460.

[4] G. Collomb and W. Härdle, *Strong uniform convergence rates in robust nonparametric time series and prediction: Kernel regression estimation from dependent observations*, Stochastic Process. Appl. **23** (1984), 77–89.

[5] D. Cox and T. Y. Kim, *Moment bounds for mixing random variables useful in nonparametric function estimation*, Stochastic Process. Appl. **56** (1995), 151–159.

[6] P. Doukhan, J. Leon, and F. Portal, *Vitesse de convergence dans la théoreme central limite pour des variables aléatoires mélangeantes a valeurs dans un espace de Hilbert*, C. R. Acad. Sci. Paris Ser. I Math. **298** (1985), 305–308.

[7] E. A. Nadaraya, *On estimating regression*, Theory Probab. Appl. **9** (1964), 141–412.

[8] P. Robinson, *Nonparametric estimators for time series*, J. Time Ser. Anal. **4** (1983), 185–207.

[9] ———, *Robust nonparametric autoregression*, Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist. **26** (1984), 247–265.

[10] G. S. Watson, *Smooth regression analysis*, Sankhyā Ser. A **26** (1964), 359–372.

Department of Statistics
Keimyung University
Taegu 704-701, Korea