Arab Journal of Sciences
and Research Publishing
Volume (8), Issue (1) : 30 Mar 2022
P: 102 - 116

ISSN: 2518-5780

المجلة العربية للعلوم
ونشـر الأبحـاث
المجلد (8)، العدد (1) : 30 مارس 2022 م
ص: 102- 116

Almost Sure Convergence of The k-NN Regression Estimate under Mixing condition

Dema Ahmad Al-shakh

Mohamed Deribati

Faculty of Science || Tishreen University || Syria

Ahmad Younso

Faculty of Science || Damascus University || Syria

Abstract: In this work we will establish almost sure convergence for k-nearest neighbor estimate of the regression function under some mixing conditions. Our results will extend some previous results in the i.i.d case to the dependent case. In addition, we will conduct a simulation study using R software program to display the importance and influence of the sample size (n) on behavior of the estimator. For this purpose, the mean squares error criterion (MSE) was used.

Keywords: Nonparametric regression, mixing concepts, k-nearest neighbor estimator, cross validation method, convergence.

التقارب شبه الأكيد لمقدّر انحدار الجوارات الـ k الأكثر قرباً تحت شرط المزج

ديمه أحمد الشاخ

محمد دريباتي

كلية العلوم || جامعة تشرين || سورية

أحمد يونسو

كلية العلوم || جامعة دمشق || سورية

المستخلص: سنقدم في هذه الورقة بعض انواع الاتساق لمقدّر الجوارات الـ k الأكثر قرباً لدالة الانحدار تحت بعض شروط المزج. نتائجنا هي توسيع لبعض الاعمال السابقة من الحالة المستقلة (i.i.d) إلى الحالة المرتبطة. بالإضافة لدراسة محاكاة باستخدام الحزمة الإحصائية **R** لمعرفة أهمية ومدى تأثير حجم العينة (n) على سلوك هذا المقدّر حيث تم استخدام معيار (MSE) متوسط مربعات الخطأ لهذا الغرض.

الكلمات المفتاحية: الانحدار اللاوسيطي، معاملات المزج، مقدّر الجوارات الـ k الأكثر قربًا، طريقة التحقق المتبادل، التقارب

Introduction.

Regression analysis is very important tools to show the relationship between variables in statistic. There are two kind of regression (parametric regression, nonparametric regression). In parametric regression we assume that we know the form of the regression function. However, the form of the true regression function is not usually known in practice, so parametric regression is not always a good choice

to estimate the regression function. So, the nonparametric regression methods are better choice to estimate the unknown regression function where no assumptions about the form of the regression function.

One of the oldest nonparametric approaches to regression analysis and pattern recognition is the nearest neighbor estimation, Fix and Hodges (1951,1952) suggested the basic idea of these nonparametric estimation rules and formalized by Royall (1966). Under the i.i.d. assumption, many results about properties of k-NN regression have been studied for a long time, and we list some of works here. The MSE, the MISE and the asymptotic normality of the k-NN regression estimate with uniform weighting function was studied by Royall (1966). Later Mack (1981) extended this result for non-uniformly weighted k-NN regression estimate. He studied the bias and the asymptotic normality under the i.i.d assumption too. Stone (1977) present the convergence for various type of k-NN estimators, strong consistency are studied by Devroye(1981), he show the almost sure convergence to 0 for special case under the boundedness of y. Devroy(1982) obtained the strong consistency and the uniform convergence. Collomb(1980) obtained other types of convergence like convergence in probability, almost surly and almost completely of the regression function estimation.

A number of works such as Mack and Silverman(1982), Cheng(1984), Devroy(1978), Li and et al.(2011), Kudraszow and View(2013) give strong uniform convergence rates. Biau et al.(2010) give guarntees under L_2 risk. Devroye et al.(1994) give consistency guarntees under the L_1 risk. We will expand results to dependent processes for the k-NN regression estimate. Clearly, we will establish the almost sure convergence of such regression estimator for $\beta - mixing$ processes.

STUDY PROBLEM:

Nonparametric regression was widely studied in the independent case, but in real application, this is not always achieved. So we want to expand the study to the dependent case.

OBJECTIVES:

We aim in this paper to extend some results in the i.i.d case to the dependent case; clearly, we aim to study almost sure convergence for k-NN regression estimator under some mixing condition.

MATERIAL AND METHODS.

1. Nearest Neighbor Regression estimate:

Let (X, Y) be a random vector defined on some probability space $(\Omega, \mathcal{A}, P)$ and talking values in $R^d \times R$. The regression function of Y given $X = x$ is defined by $m(x) = E(Y|X = x)$.

Generally $m(x)$ is unknown and one wants to estimate it using a set of copies of (X, Y). Let $\mathcal{D}_n = \{(X_1, Y_1), .., (X_n, Y_n)\}$ be a set of copies of(X, Y). Observe that (X_i, Y_i) has the same distribution of (X, Y). Many ways of estimation are suggested in the literature. One of the most simple and popular is probably the k-nearest neighbor method.

For a fixed x in R^d our goal is to estimate the regression function $m(x) = E(Y|X = x)$ using the data $\mathcal{D}_n$. Equip the space R^d with the standard Euclidean norm, and then the k-nearest neighbors estimate (k-NN) of $m(x)$ is given by

$$m_n(x) = \sum_{i=1}^{n} W_{ni}\, Y_i$$

with $W_{ni} = W_{ni}(x; X_1, .., X_n) = 1/k$ if X_i is one of the k nearest neighbors of x among $X_1, .., X_n$ and W_{ni} is 0 otherwise. Hence, $\sum_{i=1}^{n} W_{ni} = 1.$

We suppose that $k = k(n)$ with

$$k \to \infty \; as \;\; n \to \infty \qquad (1.1)$$

and

$$k/n \to 0 \; as \; n \to \infty \qquad (1.2)$$

Observe that assumptions (1.1) and (1.2) are classical to establish different kinds of consistency for the regression and density function by the k-nearest neighbor in the i.i.d case (see for example Bosq and Lecoutre(1987)).

2. Mixing conditions.

We first introduce some notations. A sequence $(Z_i, i \geq 1)$ is said to be $\alpha - mixing$ (or strongly mixing) if

$$\alpha(n) = \sup_{l \geq 1} \sup_{A \in \mathcal{F}_1^l, B \in \mathcal{F}_{n+l}^{\infty}} |P(B \cap A) - P(A)P(B)| \to 0 \; as \; n \to \infty$$

Where $\mathcal{F}_1^l, \mathcal{F}_{l+n}^{\infty}$ the sup σ −algebra generated by $(Z_i, i = 1, .., l)$ and $(Z_i, i = l + n, ..)$ respectively. The $\alpha - mixing$ coefficient is one of the most general mixing coefficient (for further details about mixing see Bradely(2005),Rio(2017) and Bosq(2012)).

It is often used to obtain asymptotic results for some estimators in nonparametric functional estimation. We suppose that

$$\alpha(n) = O(n^{-\rho}) \qquad for \; \rho > 0 \, . \qquad (2.1)$$

This means that $\alpha(n)$ tends to 0 at polynomial rate. The sequence $(Z_i, i \geq 1)$ is said to be $\beta - mixing$ if

$$\beta(n) = \sup_{l \geq 1} E\left(\sup_{A \in \mathcal{F}_1^l} |P(A) - P(A|\mathcal{F}_{n+l}^{\infty})|\right) \to 0 \; as \; n \to \infty$$

One can verify that $2\alpha(n) \leq \beta(n)$ which means that any $\beta - mixing$ sequence is $\alpha - mixing$. We suppose that

$$\beta(n) = O(n^{-\rho}) \qquad for\ \rho > 0\ . \qquad (2.2)$$

3. Preliminary Definitions and lemmas:

Definition 1: Absolutely continuous measure: A measure μ is absolutely continuous with respect to another measure λ if $\lambda(A) = 0$ implies that $\mu(A) = 0$.if a measure said to be absolutely continuous, this means absolutely continuous with respect to Lebsegues measure.

Definition 2: Consistency: One of important properties of good estimator, where we say that an estimator $\widehat{\theta_n}$ is consistent estimator of θ if

$$\lim_{n\to\infty} P\left(\left|\widehat{\theta_n} - \theta\right| < \varepsilon\right) = 1$$

i.e. the $\widehat{\theta_n}$ is consistency if, that converge to the true value of parameter being estimated as sample size increases.

The following lemmas will be used to establish the consistency results in this paper.

Lemma 3.1: Let $Z_1\ and\ Z_2$ two R −valued bounded variables. Then

$$|cov(Z_1, Z_2)| \leq 4\|Z_1\|_\infty \|Z_2\|_\infty \alpha(\sigma(Z_1), \sigma(Z_1))$$

Where $\|.\|_\infty$ denotes the supermom norm and $\sigma(Z_i)$ denotes the $\sigma - algebra$ generated by $Z_i\ for\ i = 1,2$. For the proof of lemma 3.1 we can refer to Rio(2000).

We refer the reader to Berbee(1979)for the proof of lemma (3.2):

Lemma 3.2: [Berbee's lemma]. Let Z, W be tow random variables defined up on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking their values in R^d, and let $\mathcal{A} = \sigma(Z), \mathcal{B} = \sigma(W)$. Then there exists Z^* random variable independent as Z and has the same distribution of W and satisfies $\mathbb{P}(Z \neq Z^*) = \beta(\mathcal{A}, \mathcal{B})$.

Denote $S_{x,r}$ the closed ball centered at $x \in R^d$ with radius $r > 0$.

Lemma 3.3: Let μ be an absolutely continuous probability measure on R^d. let

$$B_a(\acute{x}) = \{\acute{x}: \mu(S_{x,\|x-\acute{x}\|}) \leq a\}.$$

Then, for all $\acute{x} \in R^d$,

$$\mu(B_a(\acute{x})) \leq \gamma_d a.$$

with γ_d denotes the minimal numbers of cones centered at the origin of angle $\pi/6$ that cover R^d. depends on the dimension d only.

We refer the reader to Devroye and Györfe (1985) for the proof of Lemma 4.3 .

Lemma 3.4: [McDiarmid]. Let $X_1, .., X_n$ independent random variables taking values in a set A, and assume that $f: A^n \to R$ satisfies

$$\sup_{\substack{x_1,..,x_n \\ \acute{x}_1,..,\acute{x}_n \in A}} |f(x_1,..,x_n) - f(x_1,..,x_{i-1},\acute{x}_i,x_{i+1},..,x_n)| \leq c_i \quad 1 \leq i \leq n.$$

for some $c_1, .., c_2 > 0$

Then $\forall\ t > 0$,

$$\boldsymbol{P}\{|f(X_1, .., X_n) - Ef(X_1, .., X_n)| > t\} \leq \boldsymbol{exp}(\frac{-2t^2}{\sum_{i=1}^{n} c_i^2})$$

We refer the reader to McDiarmid (1989) for the proof of this Lemma.

4. Main results.

Denote $J_n = \int_{Rd} |m_n(x) - m(x)|\, \mu(dx)$

In this section we state the main result on the k-NN regression estimate for dependent sequence. The result established in the following theorem providing the almost sure convergence of the k-NN regression estimate for every distribution of (X, Y) with bounded Y such that $|Y| < M < \infty$ for some $M > 0$ when X has a density f. i.e. we want to prove that $J_n \to 0 \quad as\ n \to \infty$.

The above consistency result is previously investigated by Devroy et al. (1994) in the i.i.d case. The extension problem of this result to the dependent case has not been yet treated.

Theorem: Suppose that $\mathcal{D}_n$ are observations of strictly stationary $\beta - mixing$ sequence such that (2.2) with $\rho > 1$. Suppose in addition that (1.1) and (1.2)are satisfied, and

$$\frac{k}{\sqrt{n}} \longrightarrow \infty\ as\ n \longrightarrow \infty \qquad (4.1)$$

and that there exists an integer $q = q(n)\ with\ 1 \leq q \leq n/2$ such that $n \to \infty$

$$\frac{n}{qlog(n)} \to \infty \qquad (4.2)$$

and

$$\sum k^{-1} n\beta(q) < \infty \qquad (4.3)$$

Then,

$$(J_n \longrightarrow 0\ as\ n \longrightarrow \infty)\ with\ probability\ one.$$

Whereas the condition (4.1) is weaker than that of (Bosq and Lecoutre (1987), Theorem (11.3)) in i.i.d case.

Observe that if $\beta(n) = O(n^{-\rho})$ then $\alpha(n) = O(n^{-\rho})$ since $2\alpha(n) \leq \beta(n)$ with polynomial rate**.**

The strict stationary is a concept which is stronger than the identical distribution. This condition is needed to establish the strong consistency.

5. Proof:

In order to prove the theorem, we re-write $m_n(x)$ as follows:

$$m_n(x) = \frac{1}{k}\sum_{i=1}^{k} Y_{(i)}$$

Where

$$\left(X_{(1)}, Y_{(1)}\right), .., \left(X_{(n)}, Y_{(n)}\right)$$

is reordering of the data according to increasing values of $\left\|x - X_{(i)}\right\|$(ties are broken by comparing indices). Where $\|.\|$ the Euclidian norm on R^d.For fixed $\in R^d, \|x\| = (x\acute{x})^{1/2}$.

Denote $\widehat{m}_n(x) = \frac{1}{k}\sum Y_i I_{\{X_i \in S_{(x,r_n)}\}}$ *where* $r_n = r_n(x)$ satisfies that

$$\mu\left(S_{(x,r_n)}\right) = \frac{k}{n} \qquad (5.1)$$

The existence of $r_n(x)$ is ensured since X has a density f.

Proof of theorem: For the integer $\boldsymbol{q}$ defined in theorem we can write $\boldsymbol{n = 2pq + s}$ with $\boldsymbol{p\ and\ s}$ are two integers such that $< \boldsymbol{p \leq \frac{n}{2}\ and\ 0 \leq s < q}$. Without loss of generality, we suppose that $\boldsymbol{s = 0}$.

Let $\boldsymbol{Z_i = (X_i, Y_i); i = 1, .., n}$ and define the following vectors:

$$\boldsymbol{A_1 = (Z_1, .., Z_q)} \qquad \boldsymbol{B_1 = (Z_{q+1}, .., Z_{2q})}$$
$$\boldsymbol{A_2 = \left(Z_{2q+1}, .., Z_{3q}\right)} \qquad \boldsymbol{B_2 = (Z_{3q+1}, .., Z_{4q})}$$
$$\vdots \qquad\qquad \vdots$$
$$\boldsymbol{A_p = \left(Z_{2(p-1)q}, .., Z_{(2p-1)q}\right)} \qquad \boldsymbol{B_p = (Z_{(2p-1)q+1}, .., Z_{2pq})}$$

Let's define for $\boldsymbol{l = 1, .., p}$ a family of subsets in $\{\boldsymbol{1, .., n}\}$ as follows:

$$S_l = \{i: 2(l-1)q + 1 \leq i \leq (2l-1)q\}$$
$$\tilde{S}_l = \{i: (2l-1)q + 1 \leq i \leq 2lq\}$$

Observe that, for example, if $l = 1, S_l = \{1, .., q\}$ and $\tilde{S}_1 = \{q+1, .., 2q\}$.

So, $A_1 = (Z_i, i \in S_l) \quad and \quad B_1 = (Z_i, i \in \tilde{S}_1)$

Furthermore, we have $|i - j| \geq q$ for any $i \in S_l (i \in \tilde{S}_l)$ and $j \in S_{\acute{l}} (j \in \tilde{S}_{\acute{l}})$ with $l \neq \acute{l}$.

Denote:

$$A_0 = \left(Z^*{}_1, .., Z^*{}_q\right) = A_0^* \qquad , \qquad B_0 = \left(Z^*{}_{q+1}, .., Z^*{}_{2q}\right) = B_0^*$$

Using Berbee's lemma we generate $A_1^*, A_2^*, .., A_p^*$ sequence of independent vectors:

$$A_1^* = \left(Z^*{}_1, .., Z^*{}_q\right), A_2^* = (Z^*{}_{2q+1}, .., Z^*{}_{3q}), .., A_p^* = (Z^*{}_{2q+1}, .., Z^*{}_{3q})$$

Whereas A_1, A_1^* have the same probability distribution, and A_1^* is independent of A_0 with $\mathbb{P}(A_1 \neq A_1^*) \leq \beta_q$ (by Berbee's lemma), also the tow vectors A_2, A_2^* have the same probability distribution, and A_2^* is independent of A_1^*, A_1, A_0, with $\mathbb{P}(A_2^* \neq A_2) \leq \beta_q$. Thus, by the same argument on the reminder elements, we finally get that A_p^*, A_p have the same probability distribution and that A_p^* is independent of $A_0, .., A_p, A_1^*, .., A_{p-1}^*$, with $\mathbb{P}(A_p^* \neq A_p) \leq \beta_q$. In the same

way we generate a sequence of independent vectors $B_1^*,..,B_p^*$ from the sequence $B_1,..,B_p$ by using berbee's lemma whereas $\mathbb{P}(B_l^* \neq B_l) \leq \beta_q \quad \forall\, l = 1,..,p.$

Observe that the vectors $A_1^*,..,A_p^*$ are independent; and $B_1^*,..,B_p^*$ are also independent. It is easy to see that

$$\mathbb{P}(Z_i^* \neq Z_i) \leq \beta_q \quad \forall\, i = 1,..,n.$$

Or

$$\mathbb{P}((X_i^*, Y_i^*) \neq (X_i, Y_i)) \leq \beta_q \quad \forall\, i = 1,..,n.$$

But, clearly, we have:

$$\int_{R^d} |m_n(x) - m(x)|\, \mu(dx)$$

$$\leq \int_{R^d} |m_n(x) - E\hat{m}_n(x)|(dx) + \int_{R^d} |E\hat{m}_n(x) - m(x)|(dx) \qquad (5.2)$$

The second term on the right-hand side $\int_{R^d} |m(x) - E\hat{m}_n(x)|\mu(dx)$ deterministic "bias" type term, whose integral will be shown to converge to zero. According to (5.1), the condition (1.1) implies that $r_n(x) \to 0$. Note that using (5.1)

$$E\hat{m}_n(x) = \frac{n}{k} E\left(Y I_{(X \in S_{(x,r_n)})}\right)$$
$$= \frac{n}{k} E\left(E\left(Y I_{(x \in S_{(x,r_n)})} \middle| X\right)\right)$$
$$= \frac{n}{k} E\left(I_{(x \in S_{(x,r_n)})}.E(Y|X)\right)$$
$$= \frac{n}{k} \int_{S_{(x,r_n)}} E(Y|X = \acute{x})\mu(d\acute{x})$$
$$= \frac{1}{\mu(S_{(x,r_n)})} \int_{S_{(x,r_n)}} E(Y|X = \acute{x})\mu(d\acute{x})$$

So, by lebesgue's density theorem [see Wheeden and Zygmund(1977)], yields

$$E\hat{m}_n(x) = \frac{1}{\mu(S_{(x,r_n)})} \int_{S_{(x,r_n)}} E(Y|X = \acute{x})\mu(d\acute{x}) \to E(Y|X = x) = m(x).$$

Since Y is bounded, dominated convergence theorem implies that:

$$\int_{R^d} |m(x) - E\hat{m}_n(x)|\mu(dx) \to 0 \qquad (5.3)$$

Let's move to the first term in the right- hand side of (5.2). We have:

$$\int_{R^d} |m_n(x) - E\hat{m}_n(x)|\mu(dx) \leq$$

$$\int_{R^d} |\hat{m}_n(x) - E\hat{m}_n(x)|\mu(dx) + \int_{R^d} |m_n(x) - \hat{m}_n(x)|\mu(dx) \qquad (5.4)$$

Let us deal with each term in the right-hand side of (5.4). For the first term, we have

$$P\left(\int_{R^d}|\widehat{m}_n(x)-E\widehat{m}_n(x)|\mu(dx)>\frac{\varepsilon}{2}\right)\leq$$
$$P\left(\left|\int_{R^d}|\widehat{m}_n(x)-E\widehat{m}_n(x)|\mu(dx)-\int_{R^d}|\widehat{m^*}_n(x)-E\widehat{m^*}_n(x)|\mu(dx)\right|>\frac{\varepsilon}{4}\right)$$
$$+P\left(\int_{R^d}|\widehat{m^*}_n(x)-E\widehat{m^*}_n(x)|\mu(dx)>\frac{\varepsilon}{4}\right):=I+II. \qquad (5.5)$$

Where $\widehat{m^*}_n(x)=\frac{1}{k}\sum Y_i^* I_{\{X^*{}_i\in S_{(x,r_n)}\}}$.

For the term (I), for n large enough and by using Markov's inequality, we have:

$$I=P\left(\left|\int_{R^d}|\widehat{m}_n(x)-E\widehat{m}_n(x)|\mu(dx)-\int_{R^d}|\widehat{m^*}_n(x)-E\widehat{m^*}_n(x)|\mu(dx)\right|>\frac{\varepsilon}{4}\right)$$
$$I\leq\frac{E\left|\int_{R^d}|\widehat{m}_n(x)-E\widehat{m}_n(x)|\mu(dx)-\int_{R^d}|\widehat{m^*}_n(x)-E\widehat{m^*}_n(x)|\mu(dx)\right|}{\varepsilon/4}$$
$$\leq 4\varepsilon^{-1}E\int_{R^d}\left||\widehat{m}_n(x)-E\widehat{m}_n(x)|-|\widehat{m^*}_n(x)-E\widehat{m^*}_n(x)|\right|\mu(dx)$$

According to the Fubini's theorem, we get

$$I\leq 4\varepsilon^{-1}\int_{R^d}E|\widehat{m}_n(x)-\widehat{m^*}_n(x)+E\widehat{m^*}_n(x)-E\widehat{m}_n(x)|\mu(dx)$$
$$I\leq 4\varepsilon^{-1}\int_{R^d}(E|\widehat{m}_n(x)-\widehat{m^*}_n(x)|+E|\widehat{m}_n(x)-\widehat{m^*}_n(x)|)\mu(dx)$$
$$I\leq 8\varepsilon^{-1}\int_{R^d}(E|\widehat{m}_n(x)-\widehat{m^*}_n(x)|)\mu(dx)$$
$$I\leq 8\varepsilon^{-1}\int_{R^d}\left(E\left|\frac{1}{k}\sum_{i=1}^{n}Y_iI_{\{X_i\in S_{(x,r_n)}\}}-\frac{1}{k}\sum_{i=1}^{n}Y_i^*I_{\{X_i^*\in S_{(x,r_n)}\}}\right|\right)\mu(dx)$$
$$I\leq\frac{8\varepsilon^{-1}}{k}E\int_{R^d}\sum_{i=1}^{n}\left|Y_iI_{\{X_i\in S_{(x,r_n)}\}}-Y_i^*I_{\{X_i^*\in S_{(x,r_n)}\}}\right|\mu(dx)$$

Note that if $(X_i^*,Y_i^*)=(X_i,Y_i)$, then $I=0$. Generally,

$$I\leq\frac{8\varepsilon^{-1}}{k}E\left(\sum_{i=1}^{n}\int_{R^d}\left(\left|Y_iI_{\{X_i\in S_{(x,r_n)}\}}-Y_i^*I_{\{X_i^*\in S_{(x,r_n)}\}}\right|\right)\mu(dx).I_{\{(X_i^*,Y_i^*)\neq(X_i,Y_i)\}}\right)$$
$$I\leq\frac{16M\varepsilon^{-1}}{k}\sum_{i=1}^{n}P((X_i^*,Y_i^*)\neq(X_i,Y_i))$$

By the definition of $\{(X_1^*,Y_1^*),..,(X_n^*,Y_n^*)\}$ and by assumption, we have

$$I\leq\frac{16M\varepsilon^{-1}}{k}\sum_{i=1}^{n}\beta_q=\frac{16M\varepsilon^{-1}n}{k}\beta_q$$

By the assumption on k and the Borel- Cantelli lemma, we get

$$I\to 0\quad as\quad n\to\infty\quad with\ probability\ 1. \qquad (5.6)$$

Now, we move to deal with second term II, we have

$$II=P(\int_{R^d}|\widehat{m^*}_n(x)-E\widehat{m^*}_n(x)|\mu(dx)$$

$$II = P\left(\int_{R^d}\left|\frac{1}{k}\sum_{i=1}^{n} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}} - E\frac{1}{k}\sum_{i=1}^{n} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}\right|\mu(dx) > \frac{\varepsilon}{4}\right)$$

Since $n = 2pq$, we can write

$$\sum_{i=1}^{n} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}} = \sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}} + \sum_{l=1}^{p}\sum_{i\in \tilde{S}_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}$$

$$II \le P\left(\int_{R^d}\left|\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}} - E\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}\right|\mu(dx) > \frac{\varepsilon}{8}\right)$$

$$+P\left(\int_{R^d}\left|\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in \tilde{S}_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}} - E\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in \tilde{S}_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}\right|\mu(dx) > \frac{\varepsilon}{8}\right)$$

To find an upper bound for II, it suffices to treat one of the two terms in the right-hand side. Let's defined the function F:

$$F: (R^d \times R)^n \to R$$

$$F\left(A_1^*,..,A_p^*\right) = \int_{R^d}\left|\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}} - E\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}\right|\mu(dx)$$

Therefore:

$$P\left(F\left(A_1^*,..,A_p^*\right) > \frac{\varepsilon}{8}\right) \le P\left(\left|F\left(A_1^*,..,A_p^*\right) - EF\left(A_1^*,..,A_p^*\right) + EF\left(A_1^*,..,A_p^*\right)\right| > \frac{\varepsilon}{8}\right)$$

Let us prove that

$$EF\left(A_1^*,..,A_p^*\right) \to 0 \qquad (5.7)$$

to do that, denote:

$$\phi(x) = \frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}$$

Applying Cauchy-Shwartz inequality; we get, by the strict stationarity,

$$E\int_{R^d}|\phi(x) - E\phi(x)|\mu(dx) \le \int_{R^d}\sqrt{E\left(\phi(x) - E\phi(x)\right)^2}\mu(dx)$$

$$= \int_{R^d}\sqrt{var\left(\phi(x)\right)}\mu(dx)$$

$$= \int_{R^d}\sqrt{var\left(\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}\right)}\mu(dx) \qquad (5.8)$$

$$= \int_{R^d}\sqrt{\frac{1}{k^2}\sum_{l=1}^{p} var\left(\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}\right) + \frac{1}{k^2}\sum_{l=1}^{p} cov\left(\sum_{i\in S_l} Y_i^* I_{\{X_i^* \in S_{(x,r_n)}\}}, \sum_{j\in S_l} Y_j^* I_{\left\{X_j^* \in S_{(x,r_n)}\right\}}\right)}\mu(dx)$$

$$= \int_{R^d} \sqrt{\frac{1}{k^2} pq.var\left(YI_{\{X\in S_{(x,r_n)}\}}\right) + \frac{1}{k^2} p.cov\left(\sum_{i\in S_l} Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}, \sum_{j\in S_l} Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right)} \mu(dx)$$

$$var\left(YI_{\{X\in S_{(x,r_n)}\}}\right) \le M^2 \mu\left(S_{(x,r_n)}\right) \le M^2 \frac{k}{n} \qquad (5.9)$$

$$\forall\, i \neq j: \left|cov\left(Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}, Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right)\right|$$

$$\le 4 \left\|Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}\right\|_\infty \left\|Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right\|_\infty \alpha(|i-j|)$$

$$cov\left(\sum_{i\in S_l} Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}, \sum_{j\in S_l} Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right) = \sum_{\substack{i,j\in S_l \\ i\neq j}} cov\left(Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}, Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right)$$

$$\le 4 \sum_{\substack{i,j\in S_l \\ i\neq j}} \left\|Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}\right\|_\infty \left\|Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right\|_\infty \alpha(|i-j|)$$

$$\le 4M^2 \left\|I_{\{X\in S_{(x,r_n)}\}}\right\|_\infty^2 \sum_{\substack{i,j\in S_l \\ i\neq j}} \alpha(|i-j|)$$

$$\le 4M^2 q \sum_{t=1}^{\infty} \alpha\,(t)$$

By the assumption $\alpha(t) = t^{-\rho}; \rho > 1$ then, there is $c > 0$ such that

$$cov\left(\sum_{i\in S_l} Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}, \sum_{j\in S_l} Y_j^* I_{\left\{X_j^*\in S_{(x,r_n)}\right\}}\right) \le 4M^2 \sum_{t=1}^{\infty} t^{-\rho} \le 4M^2 qc$$

$$\Longrightarrow E\int_{R^d} |\phi(x) - E\phi(x)|\mu(dx) \le \int_{R^d} \sqrt{\frac{p}{k^2} qM^2 \frac{k}{n} + \frac{pq}{k^2} 4M^2 C}\,\mu(dx)$$

$$\le \sqrt{\frac{p}{k^2} qM^2 \frac{k}{n} + \frac{p}{k^2} 4qM^2 C} = \sqrt{\frac{M^2}{2k} + \frac{n}{k^2} 2M^2 C} \to 0 \qquad (5.10)$$

The last limit is a consequence of (1.2) and (4.1). By (5.8) and (5.10), we get (5.7).

Thus, for n large enough,

$$P\left(\int_{R^d} \left|\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}} - E\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in S_l} Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}\right| \mu(dx) > \frac{\varepsilon}{8}\right)$$

$$\le P\left(\left|F\left(A_1^*,..,A_p^*\right) - EF\left(A_1^*,..,A_p^*\right)\right| > \frac{\varepsilon}{16}\right)$$

Since $\left\{A_1^*,..,A_p^*\right\}$ is a sequence of independent vectors, let

$$\acute{a}_i^* = \left\{(\acute{x}_1, \acute{y}_1),..,(\acute{x}_q, \acute{y}_q)\right\}$$
$$a_i^* = \left\{(x_1, y_1),..,(x_q, y_q)\right\}$$

Then,

$$\left|F\left(a_1^*,..,a_p^*\right)-F\left(a_1^*,..,\acute{a}_i^*,a_i^*,..,a_p^*\right)\right| \leq$$
$$\int_{R^d}\left|\sum_{j=1}^{q}\left(\frac{1}{k}y_j I_{\{x_j\in S_{(x,r_n)}\}}-\frac{1}{k}\acute{y}_j I_{\{\acute{x}_j\in S_{(x,r_n)}\}}\right)\right|\mu(dx)$$

But $\left|\sum_{j=1}^{q}\left(\frac{1}{k}y_j I_{\{x_j\in S_{(x,r_n)}\}}-\frac{1}{k}\acute{y}_j I_{\{\acute{x}_j\in S_{(x,r_n)}\}}\right)\right|$ is bounded by $\frac{2qM}{k}$ and can differ from zero only if $x_j \in S_{(x,r_n)}$ or $\acute{x}_j \in S_{(x,r_n)}$. By note that $x_j \in S_{(x,r_n)}$ if and only if $\mu(S_{(x,\|x-x_j\|)}) \leq \frac{k}{n}$, and by using lemma 3.4, therefore

$$sup\int_{R^d}\left|\sum_{j=1}^{q}\left(\frac{1}{k}y_j I_{\{x_j\in S_{(x,r_n)}\}}-\frac{1}{k}\acute{y}_j I_{\{\acute{x}_j\in S_{(x,r_n)}\}}\right)\right|\mu(dx) \leq \frac{2qMk}{kn}\gamma_d = 2qM\gamma_d n^{-1}$$

By applying Mcdiarmid inequality where $i=1,..,p;\ \ c_i = 2qM\gamma_d n^{-1}$,then

$$P\left(\left|F\left(A_1^*,..,A_p^*\right)-EF\left(A_1^*,..,A_p^*\right)\right|>\frac{\varepsilon}{16}\right) \leq 2exp\left(-2\left(\frac{\varepsilon}{16}\right)^2\Big/\sum_{i=1}^{p}(2qM\gamma_d n^{-1})^2\right)$$
$$\leq 2exp\frac{-2\left(\frac{\varepsilon}{16}\right)^2}{(2pqM\gamma_d n^{-1})^2}$$
$$\leq 2exp\left(\frac{-\varepsilon^2 n}{256qM^2{\gamma_d}^2}\right)$$

In the same way we find that the second term in inequality

$$P\left(\int_{R^d}\left|\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in\tilde{S}_l}Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}-E\frac{1}{k}\sum_{l=1}^{p}\sum_{i\in\tilde{S}_l}Y_i^* I_{\{X_i^*\in S_{(x,r_n)}\}}\right|\mu(dx)>\frac{\varepsilon}{8}\right)$$
$$\leq 2exp\left(\frac{-\varepsilon^2 n}{256qM^2{\gamma_d}^2}\right)$$

As a consequence,

$$II \leq 4exp\left(\frac{-\varepsilon^2 n}{256qM^2{\gamma_d}^2}\right) \qquad (5.11)$$

Now, let us move to deal with the second term in the right-hand side of (5.4).

We have, $|m_n(x)-\hat{m}_n(x)| = \frac{1}{k}\left|\sum_{j=1}^{n}Y_j I_{\{X_j\in S_{(x,r_n)}\}}-\sum_{j=1}^{n}Y_j I_{\{X_j\in s(x,\rho_n)\}}\right|$

$$\leq \frac{M}{k}\sum_{j=1}^{n}\left|I_{\{X_j\in S_{(x,r_n)}\}}-I_{\{X_j\in S_{(x,\rho_n)}\}}\right|$$
$$= M\left|\frac{1}{k}\sum_{j=1}^{n}I_{\{X_j\in S_{(x,r_n)}\}}-1\right|$$
$$= M|\hat{g}_n(x)-E\hat{g}_n(x)|$$

Where $\hat{g}_n(x)$ is defined as $\hat{m}_n(x)$ with Y replaced by the constant random variable $\hat{Y} = 1$. Hence, by the same steps using to prove (5.9), we get,

$$\forall\, \varepsilon > 0, \quad \boldsymbol{P}\left(\int |\hat{g}_n(x) - E\hat{g}_n(x)|\mu(dx) > \varepsilon\right) \leq 4\exp\left(-\frac{cn}{q}\right) \qquad (5.12)$$

For some constant $c > 0$ depend only on d.

$$\boldsymbol{P}\{J_n \geq \varepsilon\} \leq 6\exp(-\frac{n\varepsilon^2}{256qM^2\gamma^2})$$

By talking the sum over n to both sides we get

$$\sum_{\boldsymbol{n\geq 1}} \boldsymbol{P}\{J_n \geq \varepsilon\} \leq 6\sum_{n\geq 1} \exp(-\frac{n\varepsilon^2}{256qM^2\gamma^2})$$

Since by assumption $\frac{n}{qlogn} \to \infty\ as\ n \to \infty$, by using Borel-cantelli's lemma together with (5.11) and (5.12) we get

$$II \to 0 \quad and \int |\hat{g}_n(x) - E\hat{g}_n(x)|\mu(dx) \to 0 \qquad with\ propability\ one. \quad (5.13)$$

According to (5.5),(5.6) and (5.13) we get

$$\int_{R^d} |m_n(x) - E\hat{m}_n(x)|(dx) \to 0\ with\ probability\ one. \qquad (5.14)$$

Finally, proof is completed according to (5.3) and (5.14)

6. Simulation study.

In this section we conduct a simulation study to compare the performance of the estimator for different sample sizes. To apply the k nearest neighbor method one should select an optimal k (number of neighbors) based on some criteria, for example $k = \lfloor\sqrt{n}\rfloor$ neighbors is widely used in practice, but this rule is not always feasible and may give poor results, so we propose the cross-validation criterion(CV) as smart way to select the optimal number of k.

The CV criterion is based on minimizing, with respect to k, and given by the following term

$$CV(k) = \frac{1}{n}\sum_{i=1}^{n}\left((Y_i - m_n^{-i}(x)\right)^2 w(X_i)$$

Where $m_n^{-i}(x)$ indicates the k-NN regression estimation based on leaving out the i pair (X_i, Y_i), and $w(X_i)$ is the weight function of the element X_i, we will take it as constant.

Now we use R software program to generate data (the reader can referred to Cohen(2008),Crawley (2013) and Rhys(2020) to know more information about R language). Data are simulated for βeta mixing sequence, where different sample sizes are chosen n=[50,100,200,300]. To increase robustness of the results, we generate 100 training samples of size n and 100 corresponding test samples of size 100， and average the results. We use the training sample to find the optimal k by CV(k)

criterion, and we use the Mean Squared Error (MSE) to evaluate the results based on the associated test sample.

The method is applied to the following regression model:

$$Y_i = 2 + 0.5X_i + 10\exp\left(\frac{X_i - 1}{2}\right) + \varepsilon_i$$

Where $\{\varepsilon_i\}_{i=1}^{n}$ are i.i.d random errors with mean 0 and variance 0.25.

N	5o	100	200	300
k_{opt}	K=2	K=3	K=2	K=2
AMSE	0.85817	0.662132	0.3324791	0.32971

Table 1: Estimated optimal k and average mean squared error corresponding to sample sizes.

Taple 1 shows that the estimated optimal k and the average MSE decrease when the training sample size increases. This means that the practical results in the simulation study are in line with theoretical results.

Now we try to applying k-NN regression estimator to the autoregressive models of order p . Autoregressive models AR(p) are a special case of ARMA models that are known to be β-mixing (see Mokkadaem(1988)).

Let $\{Z_t\}_{t=1}^{n}$ be a stationary time series, and $X_t = \left(Z_{t-1}, .., Z_{t-p}\right)$. AR(p) models defined as :

$$Z_t = \phi_1 Z_{t-1} + \ldots + \phi_p Z_{t-p} + \varepsilon_t \quad , \quad t = 1, \ldots, n$$

Where $\varepsilon_t \sim N(0, \sigma^2)\ and\ \phi = \left(\phi_1, .., \phi_p\right)$ is the vector of model coefficients, then regression model is: $Z_t = m(X_t) + \varepsilon_t$

For this we use R software program to generate time series of size $n = 200$ as observation of AR(1) model $z_t = 0.9Z_{t-1} + \varepsilon_t$.

First we find the optimal k value using cross validation method(k=2), then we apply k-NN estimator.

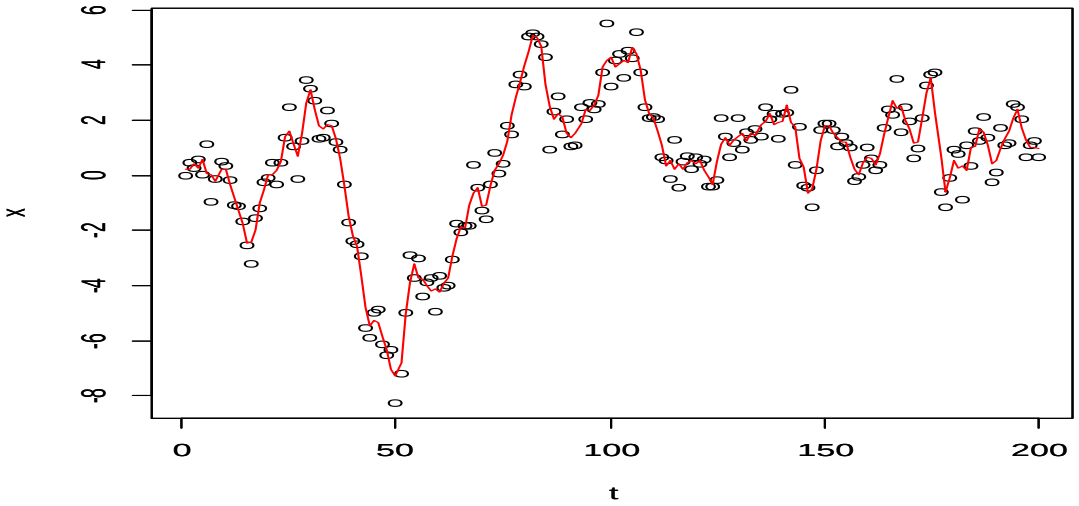

Figure (1): k-NN regression estimator to the time series with k=2.

Figure (1) shows how much the k-NN estimator is appropriate to scatter plot for time series under study.

7. Conclusion.

We see in this paper that the k-NN regression estimator converge almost surly to the regression function under β mixing condition. The simulation study showed that the quality of the estimator increased with the increase in the sample size.

We suggest expanding the study to include the case of random fields and spatial stochastic processes. We suggest too developing the results to include other types of mixing.

Reference.

- Berbee, H., Random walks with stationary increments and renewal theory, Mathematisch Centrum(1979).
- Biau, G., Gerou F., Guyader A., Rate of convergence of the Functional k-Nearest Neighbor Estimate, IEEE Transactions on information Theory 56(4), (2010)2034-2040.
- Bosq, D.; Lecoutre, J. P. Théorie de l' Estimation Fonctionnelle. Economica, Paris(1987).
- Bosq, D., Nonpaametric statistics for stochastic processes estimation and prediction, Springer Science & Business Media (2012).
- Bradley, R. C. Basic properties of strong mixing conditions. A survey and some open questions, Probab. Surv. 2(2005)107–144.
- Cheng, K. F., Strong Consistency of Nearest Neighbor Regression Function Estimators. Journal of Multivariate analysis 15, (1984)63-72.

- Cohen, J.Y. Statistics and Data with R: An Applied Approach Through Examples. A John Wiley & Sons, Ltd. (2008)618.
- -Collomb, G., Estimation de la regression la method des k points les plus proches avecnoyau, statistque non Parametrique Asymptotique, LectureNotes in Math. Spri nger Berlin821(1980)159-175.
- Crawley J. M. The R book. 2nd. ed., John Wiley & Sons, Ltd. (2013)1060.
- Devroye, L.P. Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates. Z. Wahrscheinlichkeneitstheorie Verw Gebiete 61(1982) 467-81.
- Devroye, L.P. The uniform convergence of nearest neighbor regression function estimators and their application in optimization, IEEE Trans. Inform. Theory, 24 (1978) 142–151.
- Devroye L.P.; Györfi, L. Nonparametric Density Estimation: The L1 View. Wiley:New York (1985).
- Devroye, L., Györfi, L., Krzyżak, A., and Lugosi, G. On the strong universal consistency of nearest neighbor regression function estimates. Annals of Statistics, 22 (1994)1371–1385.
- Györfi, L.; Kohler M.; Krzyzak, A.; Walk, H., A *Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York(2002)664.
- Fix, E.; Hodges, J.L, Discriminatory analysis, nonparametric discrimination, USAF School of Aviation Medicine, Randolph Field, Texas, Project 21-49-004, Report 4, Contract AD41(128)31, 1951..
- Kudraszow NL, Vieu P.Uniform consistency of kNN regressors for functional variables. Stat Probab Lett 83(8)(2013)1863-1870.
- Mack, Y.P., Local properties of k-NN Regression estimates, SIAM Journal on Algebraic and Discrete Methods 2, (1981)311-323.
- McDiarmid, c., On the Method of Bounded Differences, In Surveys in combinatories, Cambridge University Press. Cambridge (1989).
- Mokkadem, A., Mixing properties of ARMa processes, stochastic processes and their application, 29(1988)309-315.
- Rhys, I.H., Machine learning with R, The Tidyverse, and mlr, 1ed, Manning Publications(2020)1103.
- Rio, E., Asymptotic theory of weakly depended random process, . Probability Theory and Stochastic Modelling, vol 80. Springer, Berlin, Heidelberg.(2017)211
- Rio, E., Theorie Asympototique des Processus Aleatoires Faiblement Dependentes, Springer Brline(2000).
- Royall, R. M., A class of nonparameteric estimates of a smooth regression function, Ph.D. dissertation, Stanford University, Stanford(1966)68.
- Stone, C., Consistent non parametric regression, Annals of Statistics 5(1977)595-645.
- Wheeden, R.; Zygmund, A., Measure and Integral, Deeker, New York(1977)285.