Shahid Bahonar University of Kerman
Journal of Mahani Mathematical Research
Print ISSN: 2251-7952 Online ISSN: 2645-4505

# ON ALMOST SURE CONVERGENCE RATES FOR THE KERNEL ESTIMATOR OF A COVARIANCE OPERATOR UNDER NEGATIVE ASSOCIATION

H. Jabbari

*Special issue Dedicated to memory of professor Mahbanoo Tata*
Article type: Research Article
(Received: 30 December 2023, Received in revised form 15 March 2024)
(Accepted: 12 July 2024, Published Online: 07 August 2024)

Abstract. It is suppose that $\{X_n,\ n \geq 1\}$ is a strictly stationary sequence of negatively associated random variables with continuous distribution function F. The aim of this paper is to estimate the distribution of $(X_1, X_{k+1})$ for $k \in \mathbb{N}_0$ using kernel type estimators. We also estimate the covariance function of the limit empirical process induced by the sequence $\{X_n,\ n \geq 1\}$. Then, we obtain uniform strong convergence rates for the kernel estimator of the distribution function of $(X_1, X_{k+1})$. These rates, which do not require any condition on the covariance structure of the variables, were not already found. Furthermore, we show that the covariance function of the limit empirical process based on kernel type estimators has uniform strong convergence rates assuming a convenient decrease rate of covariances $Cov(X_1, X_{n+1}),\ n \geq 1$. Finally, the convergence rates obtained here are empirically compared with corresponding results already achieved by some authors.

*Keywords*: Almost sure convergence rate, Bivariate distribution function, Empirical process, Kernel estimation.
*2020 MSC*: Primary 62G20, 60F15, 62G05.

## 1. Introduction

Estimation of distribution functions of random pairs has been always a subject of interest for many statisticians. It was studied by several authors under the assumption that the random variables are independent (see, for example, [4]). However, it is not always reasonable to consider the concerned random variables as independent. The case of nonindependent random variables has been studied, too (see, for example, [1], [2], [9], [10], and [11]). Negative association (NA) is one of the most applicable concepts of negative dependence in multivariate statistical analysis and reliability theory ( [19]). Because the NA sequence includes the independent sequence, it has been widely applied in multivariate statistical analysis, the permeability analysis, and reliability theory

✉ jabbarinh@um.ac.ir, ORCID: 0000-0002-7028-9052
https://doi.org/10.22103/jmmr.2024.22733.1561
Publisher: Shahid Bahonar University of Kerman

How to cite: H. Jabbari, *On almost sure convergence rates for the kernel estimator of a covariance operator under negative association*, J. Mahani Math. Res. 2024; 13(3): 91 - 104.

drawn much attention, and a lot of research results has been obtained. Some cases of negatively associated random variables are normal random variables (with negative correlation), multinomial, convolution of unlike multinomial, multivariate hypergeometric, Dirichlet, and Dirichlet compound multinomial distributions (see [15]).

A finite family of random variables $\{X_i,\ 1 \leq i \leq n\}$ is said to be negatively associated (NA) if for every pair of disjoint subsets $A$ and $B$ of $\{1, 2, \ldots, n\}$,

$$Cov(f_1(X_i, i \in A), f_2(X_j, j \in B)) \leq 0,$$

whenever $f_1$ and $f_2$ are coordinatewise increasing and such that the covariance exists. An infinite family of random variables is NA if every finite subfamily is NA. It is clear that sequences of NA random variables are a family of very wide scope, which contain independent random variables. We refer the reader to [19], [1], [15], [29], [24], [23], [22], [27], [13], [11], [14], [25], [28], [12], [18] and [26] for more results on different aspects of NA random variables.

The mentioned comments above have drawn attention to the estimation of the bivariate distribution function under negative association. A natural (histogram) estimator of $F_k(r, s) = P(X_1 \leq r, X_{k+1} \leq s)$ with $k$ fixed, is defined by

$$\tilde{F}_k(r, s) = \frac{1}{n-k} \sum_{i=1}^{n-k} \{1_{(-\infty, r]}(X_i) 1_{(-\infty, s]}(X_{k+i})\}, \tag{1}$$

where $1_A$ denotes the indicator of event $A$. The asymptotic behavior of this estimator was studied by [9], [10] and [13]. For dependent sequences, under certain conditions (see Theorem 17 and the first remark in page 137 of [20]), the limit of the uniform empirical process still is a centered Gaussian process, but the covariance function changes to

$$\Gamma_k(r, s) = \varphi_k(r, s) + \sum_{k=1}^{\infty} \varphi_k(r, s) + \sum_{k=1}^{\infty} \varphi_k(s, r),$$

where $\varphi_k(r, s) = F_k(r, s) - F(r)F(s)$. Under a convenient decrease rate of the covariances, [9], [10] and [13] obtained a uniform strong convergence rate of $n^{-1/2}$ for two-dimensional empirical distribution function of $(X_1, X_{k+1})$ and covariance function of the limit empirical process.

Kernel estimators of density and distribution functions are the most well-known nonparametric methods that their properties have been studied and understood for decades by many authors (see, for example, [16], [3], [8] and [7]). [2] and [11] considered the kernel estimator of $F_k$, defined by

$$\hat{F}_k(r, s) = \frac{1}{n-k} \sum_{i=1}^{n-k} U\left(\frac{r - X_i}{h_n}, \frac{s - X_{k+i}}{h_n}\right), \tag{2}$$

where $U(., .)$ is a given bivariate distribution function and $\{h_n, n \geq 1\}$ is a sequence of positive numbers converging to zero. They found the optimal bandwidth convergence rate of order $n^{-1}$. In this paper, using $\hat{F}_k$ in (2), we define the kernel estimator of $\varphi_k(r,s)$ and $\Gamma(r,s)$ as

$$\hat{\varphi}_k(r,s) = \hat{F}_k(r,s) - \hat{F}(r)\hat{F}(s),\ \hat{\Gamma}(r,s) = \hat{\varphi}_k(r,s) + \sum_{k=1}^{n}(\hat{\varphi}_k(r,s) + \hat{\varphi}_k(s,r))$$

and derive a uniform convergence rate of order $h_{n-k}^2 n^{-\gamma}$ for the above estimators, where $0 < \gamma < 1/2$ and $\hat{F}(r)$ is obtained from $\hat{F}_k(r,s)$ with $k = 0$ and $s = r$. In fact $\hat{F}(r) = \frac{1}{n}\sum_{i=1}^{n} U(\frac{r-X_i}{h_n})$, where $U(.)$ is a univariate kernel distribution function which is the special case of $U(.,.)$. For the convergence rate, we need no condition on the covariance structure of the variables. The above rate is flexible because of including the term $h_n$ which can be optionally chosen. This flexibility makes us able to have a rate that tends to zero (as is necessary for a convergence rate) and on the other hand, can be a better rate than what was found by [13] and [11].

In what follows, we suppose that $C$ is a positive constant not depending on $n$. Also, we use the following general assumption throughout the paper:

**(A)**. $\{X_n,\ n \geq 1\}$ is a NA and strictly stationary sequence of random variables having bounded density function and

$$|U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) - EU(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n})| \ \leq\ Ch_n^2,\ a.s. \tag{3}$$

for any $1 \leq i \leq n$ and fixed $r, s \in I\!R$.

*Remark* 1.1. It can be easily checked that (3) holds for any NA sequence of random variables mentioned in $(A)$, because

$$\begin{aligned} U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) \ &-\ EU(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) = \int_{-\infty}^{Y_2}\int_{-\infty}^{Y_1} u(t_1,t_2)dt_1dt_2 \\ &-\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(\frac{r-x_i}{h_n}, \frac{s-x_{i+k}}{h_n})dF_k(x_i, x_{i+k}),\ a.s. \end{aligned}$$

where $Y_1 = \frac{r-X_i}{h_n}$, $Y_2 = \frac{s-X_{i+k}}{h_n}$, and $u(.,.)$ is the probability density function associated to $U(.,.)$. By taking $t_1 = r + h_n z_1$, $t_2 = s + h_n z_2$, $x_i = r - h_n v_i$ and $x_{i+k} = s - h_n v_{i+k}$ in the above equation and some calculations, we'll have

$$\begin{aligned} &|U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) - EU(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n})| \\ &\leq\ h_n^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(r+h_n z_1, s+h_n z_2)dz_1dz_2 \\ &\quad +\ h_n^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(v_i, v_{i+k})dF_k(v_i, v_{i+k}) \ =\ O(h_n^2).\ a.s. \end{aligned}$$

In Section 2, some auxiliary results are given to find the convergence rates. The moment inequality used for the proofs is presented in this section. The strong uniform convergence rates are proved in Sections 3 and 4. In Section 5, we compare the histogram and kernel estimators graphically using statistical computations and then conclude the results.

## 2. **Auxiliary results**

In this section, we use the following moment inequality for NA random variables and obtain an essential inequality required to prove our convergence rates.

**Lemma 2.1.** *[ [17] and [24]] Let $(X_1, X_2, \ldots, X_n)$ be an NA random vector with $EX_j = 0$ and $E|X_j|^p < \infty$ for some $p \geq 2$ and all $j = 1, \ldots, n$. Then, there exists a constant $C = C(p) > 0$ such that*

$$E|\sum_{j=1}^{n} X_j|^p \quad \leq \quad C[\sum_{j=1}^{n} E|X_j|^p + (\sum_{j=1}^{n} EX_j^2)^{p/2}].$$

**Lemma 2.2.** *Let $k \in \mathbb{N}_0$ be fixed, and let $\varepsilon_n, n \geq 1$ be a sequence of positive numbers. Suppose that the assumption $(A)$ is satisfied. Then, there exists a constant $C$ such that, for $r, s \in \mathbb{R}$ and $p > 2$,*

$$P(|\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) \quad \leq \quad \frac{Ch_{n-k}^{2p}}{\varepsilon_n^p (n-k)^{p/2}}.$$

**Proof**. Similar to the technique used for the proof of Lemma 2.2 in [13], for each $n \in \mathbb{N}$, $1 \leq i \leq n$ and fixed $r, s \in \mathbb{R}$, take

$$Z_{k,i} \quad = \quad U(\frac{r - X_i}{h_n}, \frac{s - X_{i+k}}{h_n}) - F_k(r,s)$$

and also

$$W_{k,i} \quad = \quad Z_{k,i} - E(Z_{k,i}).$$

So, we have

$$\begin{aligned}\hat{F}_k(r,s) - E(\hat{F}_k(r,s)) &= \frac{1}{n-k}\sum_{i=1}^{n-k} Z_{k,i} + F_k(r,s) - E(\hat{F}_k(r,s)) \\ &= \frac{1}{n-k}\sum_{i=1}^{n-k} W_{k,i} + \frac{1}{n-k}\sum_{i=1}^{n-k} E(Z_{k,i}) + F_k(r,s) - E(\hat{F}_k(r,s)).\end{aligned}$$

Since $\frac{1}{n-k}\sum_{i=1}^{n-k} E(Z_{k,i}) = E(\hat{F}_k(r,s)) - E(F_k(r,s))$, we have

$$\hat{F}_k(r,s) - E(\hat{F}_k(r,s)) \quad = \quad \frac{1}{n-k}\sum_{i=1}^{n-k} W_{k,i}.$$

Under (A), it is clear that for given $k$ and $n$, $W_{k,n}$ is a decreasing function of the variables $X_n$. So, according to the properties of NA random variables (see, for more information, [15]), $\{W_{k,n},\ n \geq 1\}$ is NA and strictly stationary. Also, $|W_{k,n}| \leq Ch_n^2$ and $E(W_{k,n}) = 0$; then, $E|W_{k,n}|^p < \infty$, for each $n \geq 1$ and

$p > 2$, and so we can apply Lemma 2.1 to the sequence $\{W_{k,n},\ n \geq 1\}$. Thus, for all $n \geq 1$, we obtain

$$\begin{aligned} E|\sum_{i=1}^{n} W_{k,i}|^p &\leq C[\sum_{i=1}^{n} E|W_{k,i}|^p + (\sum_{i=1}^{n} EW_{k,i}^2)^{p/2}] \\ &\leq Cn^{p/2}h_n^{2p}. \end{aligned}$$

Now, for fixed $r, s \in I\!R$, we find, for all $n > k$,

$$\begin{aligned} P(|\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) &\leq \frac{2^p}{\varepsilon_n^p (n-k)^p} E|\sum_{i=1}^{n-k} W_{k,i}|^p \\ &\leq \frac{Ch_{n-k}^{2p}}{\varepsilon_n^p (n-k)^{p/2}}. \ \square \end{aligned}$$

As considered in [13], to prove the main results, we recall the following notations as introduced in [13]. Let $\{t_n,\ n \geq 1\}$ be a sequence of positive integers such that $t_n \to +\infty$ and for $n \in I\!N$ and $i = 1, \ldots, t_n$, put $x_{n,i} = Q(i/t_n)$, where $Q$ is the quantile function of $F$. Also, for $n \in I\!N$ and $k \in I\!N_0$, take

$$D_{n,k} = \sup_{r,s \in I\!R} |\hat{F}_k(r,s) - F_k(r,s)|,\ D_{n,k}^* = \max_{i,j=1,\ldots,t_n} |\hat{F}_k(x_{n,i}, x_{n,j}) - F_k(x_{n,i}, x_{n,j})|.$$

Now, following the same steps as in Lemma 2.4 of [13] and applying Lemma 2.2, Theorem 2 of [9] and Lemma 2.3 of [13], we can prove the next lemma.

**Lemma 2.3.** *Let $\varepsilon_n$ and $t_n$ be two sequences of positive numbers such that $t_n \to +\infty$ and $\varepsilon_n t_n \to +\infty$, and let $p > 2$ and $k \in I\!N_0$ be fixed. Suppose that $(A)$ holds. Then, for any large enough $n$,*

$$P(\sup_{r,s \in I\!R} |\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) \leq \frac{Ct_n^2}{\varepsilon_n^p (n-k)^{p/2}} h_{n-k}^{2p}.$$

## 3. Uniform strong convergence rates of $\hat{F}_k$

In this section, we summarize the previous results to get uniform strong convergence rates of $\hat{F}_k$.

**Lemma 3.1.** *Let $k \in I\!N_0$ be fixed, and suppose that $(A)$ holds. Then, under the conditions of Lemma 2.3 and for every $0 < \delta < \frac{p-2}{2}$, we have*

$$\sup_{r,s \in I\!R} |\hat{F}_k(r,s) - F_k(r,s)| = O(h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{p-2-2\delta}{2(p+2)}}) \ a.s. .$$

**Proof**. Put $t_n = \frac{1}{\varepsilon_n h_{n-k}}$, and let $0 < \delta < \frac{p-2}{2}$. Since $t_n \to \infty$ and $t_n \varepsilon_n \to \infty$ when $n \to \infty$, from Lemma 2.3 for $\varepsilon_n = h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{p-2-2\delta}{2(p+2)}}$ and $n$ large enough, we obtain

$$P(\sup_{r,s \in I\!R} |\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) \leq \frac{C}{\varepsilon_n^{p+2} h_{n-k}^{2-2p} (n-k)^{p/2}} \leq Cn^{-(1+\delta)}. \tag{4}$$

The proof is completed by using the Borel–Cantelli Lemma, because for all $\delta > 0$, the sequence on the right-hand side of (4) is summable . □

If $p \to \infty$, then $\varepsilon_n \to h_{n-k}^2 n^{-1/2}$. Since $h_{n-k}^2 \to 0$ when $n \to \infty$, the convergence rate of Lemma 3.1 remains reasonable for a large $p$. So, using Lemma 3.1 and some calculations, we summarize the results of this section in the next theorem.

**Theorem 3.2.** *Under the assumptions of Lemma 3.1 and for every* $0 < \gamma < 1/2$*, we have*

$$\sup_{r,s \in I\!R} |\hat{F}_k(r,s) - F_k(r,s)| \quad = \quad O(h_{n-k}^2 n^{-\gamma}) \ a.s. \ .$$

*Remark* 3.3. Note that Theorem 4 of [11] holds for $\hat{F}_k$ defined in (2) under some regularity assumptions. So, for all $x, y \in I\!R$, we have

$$(n-k)MSE[\hat{F}_n(x,y)] \quad = \quad F(x,y) - F^2(x,y) + 2\sum_{j=2}^{\infty}(F_j(x,y,x,y) - F^2(x,y)) + O(h_n + nh_n^2) + a_n,$$

where, for each positive integer $j$, $F_j$ is the distribution function of $(X_1, X_{k+1}, X_j, X_{k+j})$ and

$$a_n \quad = \quad \frac{1}{(n-k)}\sum_{j=2}^{\infty}(j-1)(F_j(x,y,x,y) - F^2(x,y)) - 2\sum_{j=n-k-1}^{\infty}(F_j(x,y,x,y) - F^2(x,y)).$$

Then, an optimal convergence rate of the MSE is achieved by choosing $h_n = Cn^{-1}$.

As already said, if $k = 0$ and $s = r$, then the estimator $\hat{F}_k(r,s)$ becomes to the one-dimensional kernel distribution function $\hat{F}(r)$. So, the results of Theorem 3.2 hold true for $\hat{F}$ and we can write

$$\sup_{r \in I\!R} |\hat{F}(r) - F(r)| \quad = \quad O(h_n^2 n^{-\gamma}) \ a.s. \ .$$

*Remark* 3.4. From the results of Theorem 3.2, we understand that the convergence rate $h_{n-k}^2 n^{-\gamma}$, for every $0 < \gamma < 1/2$ and $h_n$, is very faster than those obtained later by [13] (i.e. $n^{-\gamma}$). So, the kernel estimator of two-dimensional and one-dimensional distribution function $F_k$ and $F$ is better than empirical one, respectively.

Now, we can readily obtain the convergence rate of the kernel estimator of $\varphi_k$ in the next theorem and then the proof is omitted.

**Theorem 3.5.** *Under the assumptions of Theorem 3.2 and for every* $0 < \gamma < 1/2$*, we have*

$$\sup_{r,s \in I\!R} |\hat{\varphi}_k(r,s) - \varphi_k(r,s)| \quad = \quad O(h_{n-k}^2 n^{-\gamma}) \ a.s. \ .$$

## 4. Uniform strong convergence rates of $\hat{\Gamma}$

In this section, we prove the uniform strong convergence rate for the sum $\sum_{k=1}^{\infty}\hat{\varphi}_k(r,s)$, which is sufficient to obtain the desired result for the kernel estimator of $\Gamma$. For this, we need a regular assumption on the covariance structure of NA random variables. Regarding that this covariance structure highly determines its approximate independence (see [20]), it is common to have an assumption on the covariance structure of the random variables.

Now, we consider the following notation

$$v(n) = \sum_{j=n+1}^{\infty} |Cov(X_1, X_j)|^{1/3}, \tag{5}$$

to prove the uniform strong convergence rate for the kernel estimator of $\Gamma$.

**Lemma 4.1.** *Let (A) hold, $\theta > 0$ and $a_n = n^{\frac{p-2-2\delta}{p^2+3p}}$ for some $p > 2$ and each $0 < \delta < \frac{p-2}{2}$. If*

$$v(a_n) \leq C h_{n-k}^{\frac{4\theta(p-1)}{(p-2)(p+3)}} a_n^{-\theta}, \tag{6}$$

*for all $n \geq 1$, then*

$$\sup_{r,s\in I\!R} |\sum_{k=1}^{a_n}\hat{\varphi}_k(r,s) - \sum_{k=1}^{\infty}\varphi_k(r,s)| = O(h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}})\ a.s.\ . \tag{7}$$

**Proof**. The idea is essentially the same as the proof of Lemma 4.1 of [13]. So, we repeat their proof using our required notations and definitions. Put $\varepsilon_n = h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}$, for each $0 < \delta < \frac{p-2}{2}$, and $t_n = \frac{a_n}{\varepsilon_n h_{n-k}}$. Now, we may have

$$P(\sup_{r,s\in I\!R} |\sum_{k=1}^{a_n}(\hat{F}_k(r,s) - F_k(r,s))| > \varepsilon_n) \leq \sum_{k=1}^{a_n} P(\sup_{r,s\in I\!R} |\hat{F}_k(r,s) - F_k(r,s)| > \frac{\varepsilon_n}{a_n}). \tag{8}$$

Since $0 < \delta < \frac{p-2}{2}$, $\frac{(p-2)(p-2-2\delta)}{2p(p+2)} > 0$ and $0 < \frac{p-2-2\delta}{p^2+3p} < 1$, it is easy to see that $\varepsilon_n \to 0$, $a_n \to +\infty$, $t_n \to +\infty$, $\frac{\varepsilon_n}{a_n} t_n \to +\infty$, and $\frac{a_n}{n} \to 0$ as $n \to +\infty$.

Using $\frac{\varepsilon_n}{a_n}$ in place of $\varepsilon_n$ in Lemma 2.3, we obtain, for all $n$ large enough,

$$\begin{aligned} P(\sup_{r,s\in I\!R} |\sum_{k=1}^{a_n}(\hat{F}_k(r,s) - F_k(r,s))| > \varepsilon_n) &\leq \sum_{k=1}^{a_n} \frac{C t_n^2 a_n^p}{\varepsilon_n^p (n-k)^{p/2}} h_{n-k}^{2p} \\ &\leq \frac{C a_n^{p+3}}{\varepsilon_n^{p+2}(n-a_n)^{p/2}} h_{n-k}^{2p-2}. \end{aligned} \tag{9}$$

By some calculations, we may write $\varepsilon_n = h_{n-k}^{\frac{2p-2}{p+2}} a_n^{\frac{p+3}{p+2}} n^{-\frac{p-2-2\delta}{2p+4}}$. Inserting this on the right-hand side of (9) leads to summable upper bound as $\frac{a_n}{n} \to 0$. So, we have, by Borel–Cantelli Lemma,

$$\sup_{r,s\in I\!R}|\sum_{k=1}^{a_n}(\hat{F}_k(r,s)-F_k(r,s))| \quad = \quad O(h_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}})\ a.s.\ . \tag{10}$$

Now, we have

$$\begin{aligned}\sup_{r,s\in I\!R}|\sum_{k=1}^{a_n}\hat{\varphi}_k(r,s)-\sum_{k=1}^{\infty}\varphi_k(r,s)| \quad \leq \quad & \sup_{r,s\in I\!R}|\sum_{k=1}^{a_n}(\hat{F}_k(r,s)-F_k(r,s))| \\ & + 2a_n\sup_{r\in I\!R}|\hat{F}(r)-F(r)| \\ & + \sup_{r,s\in I\!R}|\sum_{k=a_n+1}^{\infty}\varphi_k(r,s)|.\end{aligned} \tag{11}$$

For the first term on the right-hand side of (11), we use (10). Since $\frac{p+3}{p+2}>1$ by using Lemma 3.1 for the second term, we have

$$\begin{aligned}a_n\sup_{r\in I\!R}|\hat{F}(r)-F(r)| \quad &= \quad O(a_nh_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{p-2-2\delta}{2p+4}}) \\ &= \quad O(h_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}})\ a.s.\ .\end{aligned}$$

For the third term on the right-hand side of (11), we use corollary of Theorem 1 in [21] and the relation (21) in [19]. So by (6) for $\theta=\frac{(p-2)(p+3)}{2p+4}>0$ and $a_n=n^{\frac{p-2-2\delta}{p^2+3p}}$, we obtain

$$\begin{aligned}\sup_{r,s\in I\!R}|\sum_{k=a_n+1}^{\infty}\varphi_k(r,s)| \quad &\leq \quad C\sum_{k=a_n+1}^{\infty}|Cov^{1/3}(X_1,X_{k+1})| \\ &= \quad Cv(a_n) \\ &= \quad Ch_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}.\end{aligned}$$

Hence, the proof is completed. □

We now summarize the above result in the following theorem.

**Theorem 4.2.** *Under the assumptions of Lemma 4.1 and condition (6), for all $n\geq 1$, $\theta>0$, and $0<\gamma<1/2$, we have*

$$\sup_{r,s\in I\!R}|\sum_{k=1}^{a_n}\hat{\varphi}_k(r,s)-\sum_{k=1}^{\infty}\varphi_k(r,s)| \quad = \quad O(h_{n-k}^2n^{-\gamma})\ a.s.\ .$$

**Proof**. For each $\delta>0$ and $p>2$, we take $\frac{(p-2)(p-2-2\delta)}{2p(p+2)}>\gamma$. So, similar to the proof of Theorem 4.1 of [13], we obtain the desired result. □

Now, summarizing the results achieved in Theorems 3.5 and 4.2, we can state the following theorem for $\hat{\Gamma}$.

**Theorem 4.3.** *Suppose that (A) holds. Under condition (6) for all $n \geq 1$, $\theta > 0$, $p > 2$, and $0 < \gamma < 1/2$, we get*

$$\sup_{r,s\in\mathbb{R}} |\hat{\Gamma}(r,s) - \Gamma(r,s)| \quad = \quad O(h_{n-k}^2 n^{-\gamma}) \ a.s. \ .$$

*Remark* 4.4. As stated in Remark 3.4, our convergence rate $h_{n-k}^2 n^{-\gamma}$, for every $0 < \gamma < 1/2$ and $h_n$ in Theorem 4.3, is very faster than those obtained by [13] (i.e. $n^{-\gamma}$). So, the kernel estimator of $\Gamma$ is better than the empirical one.

## 5. Statistical computations

In this section, we intend to compare the behavior of our estimator with those of [13] via statistical computations using R software. As noted in [5], [**?**] and [15], a number of well known multivariate distributions such as multivariate normal distribution with negative correlations possess the NA property. So for generating the NA data, suppose that the random vector $(X_1, \ldots, X_n)'$ has multivariate normal distribution with zero mean vector and covariance matrix

$$\Sigma_n = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & -\rho^2 & \cdots & -\rho^{n-1} \\ -\rho & 1 & -\rho & \cdots & -\rho^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\rho^{n-1} & -\rho^{n-2} & -\rho^{n-3} & \cdots & 1 \end{bmatrix}, \tag{12}$$

where $\rho > 0$. For $n = 20, 100$, we generate one sample from $n$-dimensional

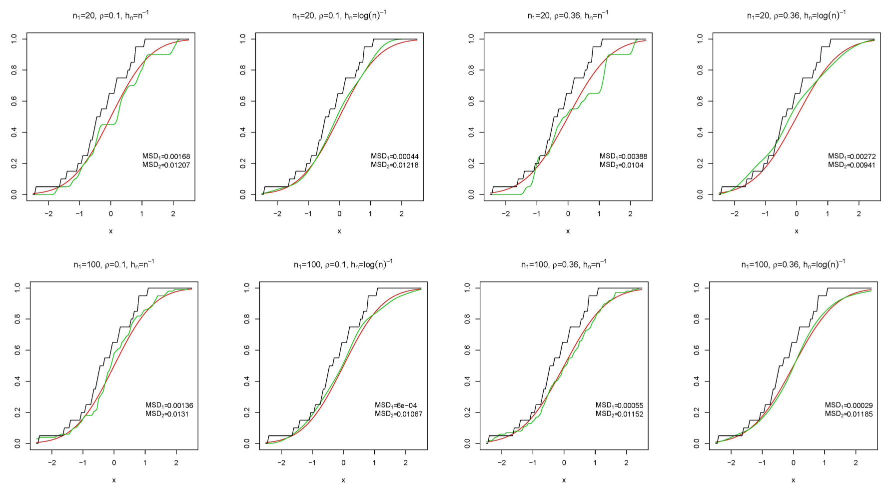


FIGURE 1. One-dimensional normal distribution function with zero mean and variance $\frac{1}{1-\rho^2}$ (red), histogram estimator (black) and kernel estimator (green) of $F(r)$.

multivariate normal distribution with zero mean vector and covariance matrix $\Sigma_n$ assuming $\rho = 0.1, 0.36$. Then, for $k = 0, 1, 2$, we compute the histogram estimator $\tilde{F}_k$ in (1) and the kernel estimator $\hat{F}_k$ in (2) using $h_n = n^{-1}$ and $h_n = \log^{-1}(n)$ and $U(.,.)$ as bivariate normal distribution with zero mean vector and covariance matrix

$$\frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}. \tag{13}$$

*Remark* 5.1. Notice that, it is important that to choose the values of $\rho$ such that **(A)** holds. Numerical results indicate that if $\rho \in [0, 0.36]$, then $|\Sigma_n|$ in (12) is positive. Due to the similarity of the results, only the cases $\rho = 0.1$ (weak dependence) and $\rho = 0.36$ (strong dependence) are reported here. Also, the simulation study is carried out for the reasonable special cases $h_n = n^{-1}$ and $h_n = \log^{-1}(n)$ which are suitable for justifying the results.

Results for $k = 0, 1, 2$ and different values of $n$, $\rho$ and $h_n$ are presented in Figures 1–3, respectively. Also for simple comparison, we compute the following mean square deviations (MSDs) between $F_k(r,s)$ and $\hat{F}_k(r,s)$ (or $\tilde{F}_k(r,s)$) for all $r, s$:

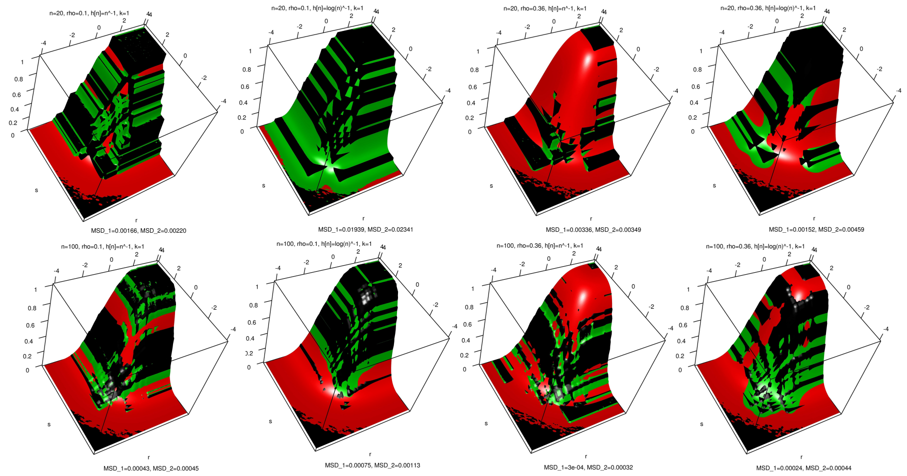


FIGURE 2. Bivariate normal distribution function with zero mean vector and covarince matrix (13) (red), histogram estimator (black) and kernel estimator (green) of $F_1(r,s)$.

$$\begin{aligned} MSD_1 &= \frac{1}{N}\sum_{r,s}(\hat{F}_k(r,s) - F_k(r,s))^2 \\ MSD_2 &= \frac{1}{N}\sum_{r,s}(\tilde{F}_k(r,s) - F_k(r,s))^2, \end{aligned} \tag{14}$$

where $N$ is the product of all numbers $r$ and $s$. Note that the values of MSDs are shown in the figures. Furthermore, for best comprehension, these values for $k = 1, 2$ in two-dimensional cases are presented in Table 1.

From Figure 1, it is observed that for $k = 0$ (one-dimensional distribution function):

**a)** When $n$ is small ($n = 20$) and large ($n = 100$), kernel estimator (green) of $F(r)$ is better than histogram estimator (black) for all values of $\rho$ and bandwidth rates $h_n$.

**b)** When $n$ becomes large, the kernel estimator has a good fit.

**c)** When $n$ is small, the bandwidth rate $h_n = \log^{-1}(n)$ is better than $h_n = n^{-1}$.

**d)** When $n$ is large, the bandwidth rates $h_n = n^{-1}$ and $h_n = \log^{-1}(n)$ have the same behaviors.

**e)** Since the kernel estimator is smooth, the best estimator of $F(r)$ is the kernel estimator.

**f)** In all graphs, MSD of kernel estimator is less than histogram estimator.

**g)** In all cases, the histogram estimator has an over estimate.

Table 1. Values of MSDs for bivariate cases and different values of $k, n, \rho$ and $h_n$.

| $k$ | $n$ | $\rho$ | $h_n$ | $MSD_1$ | $MSD_2$ |
|---|---|---|---|---|---|
| 1 | 20 | 0.1 | $n^{-1}$ | **0.00166** | 0.00220 |
| 1 | 20 | 0.36 | $n^{-1}$ | **0.00336** | 0.00349 |
| 1 | 20 | 0.1 | $\log^{-1}(n)$ | **0.01939** | 0.02341 |
| 1 | 20 | 0.36 | $\log^{-1}(n)$ | **0.00152** | 0.00459 |
| 1 | 100 | 0.1 | $n^{-1}$ | **0.00043** | 0.00045 |
| 1 | 100 | 0.36 | $n^{-1}$ | **0.00030** | 0.00032 |
| 1 | 100 | 0.1 | $\log^{-1}(n)$ | **0.00075** | 0.00113 |
| 1 | 100 | 0.36 | $\log^{-1}(n)$ | **0.00024** | 0.00044 |
| 2 | 20 | 0.1 | $n^{-1}$ | **0.00556** | 0.00621 |
| 2 | 20 | 0.36 | $n^{-1}$ | **0.00176** | 0.00238 |
| 2 | 20 | 0.1 | $\log^{-1}(n)$ | **0.00070** | 0.00190 |
| 2 | 20 | 0.36 | $\log^{-1}(n)$ | **0.00029** | 0.00167 |
| 2 | 100 | 0.1 | $n^{-1}$ | **0.00126** | 0.00129 |
| 2 | 100 | 0.36 | $n^{-1}$ | **0.00196** | 0.00207 |
| 2 | 100 | 0.1 | $\log^{-1}(n)$ | **0.00016** | 0.00023 |
| 2 | 100 | 0.36 | $\log^{-1}(n)$ | **0.00011** | 0.00044 |

Figure 2 and Table 1 indicate that, for $k = 1$ (two-dimensional distribution function with lag one):

**a)** When $n$ is small ($n = 20$), we have over estimate for weak dependence ($\rho = 0.1$) and $h_n = \log^{-1}(n)$. Also, this wrong fit holds true when $n$ is small ($n = 20$), $\rho = 0.1$, and $h_n = n^{-1}$ for some values of $r$ and $s$ (that is $r, s \in [-2, 4]$, approximately).

**b)** MSD of kernel estimator is less than histogram estimator for all cases.

**c)** When $n$ is large ($n = 100$), the difference between kernel and histogram estimators is very small.

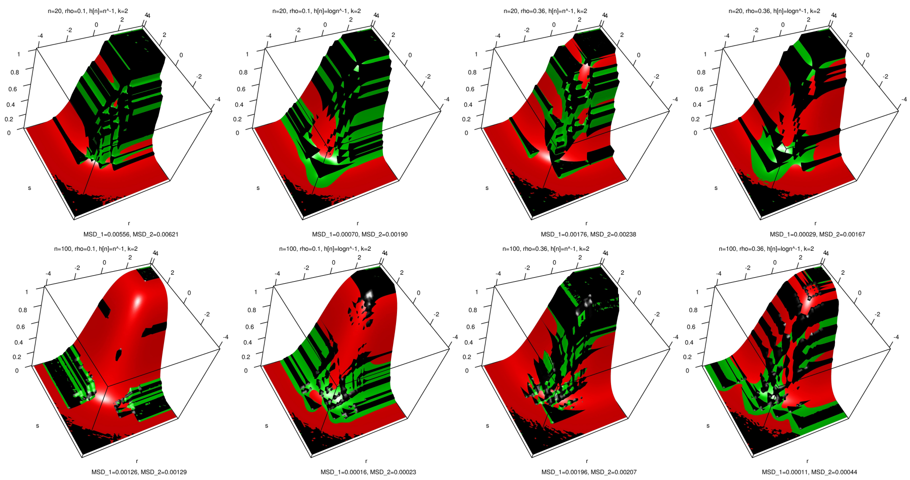


FIGURE 3. Bivariate normal distribution function with zero mean vector and covarince matrix (13) (red), histogram estimator (black), and kernel estimator (green) of $F_2(r, s)$.

**d)** When $n$ is small ($n = 20$) or large ($n = 100$), the bandwidth rate $h_n = n^{-1}$ has a better role than $h_n = \log^{-1}(n)$ for estimating $F_1(r, s)$ in weak ($\rho = 0.1$) dependence case and in strong ($\rho = 0.36$) dependence case; the bandwidth rate $h_n = \log^{-1}(n)$ is almost better than $h_n = n^{-1}$ for estimating $F_1(r, s)$.

The following results are obtained from Figure 3 and Table 1 for $k = 2$ (two-dimensional distribution function with lag two):

**a)** When $n$ is small ($n = 20$) and $\rho = 0.1$, we have over estimate for large values of $r$ and $s$ (that is, $r, s \in [0, 4]$, approximately).

**b)** MSD of kernel estimator is less than histogram estimator for all cases.

**c)** When $n$ is large ($n = 100$), the difference between kernel and histogram estimators is very small.

**d)** When $n$ is small ($n = 20$) or large ($n = 100$), the bandwidth rate $h_n = \log^{-1}(n)$ has a better role than $h_n = n^{-1}$ for estimating $F_2(r, s)$, approximately.

## 6. Aknowledgement

I am grateful to the editor and two referees whose insightful and constructive comments have led to improvement of the manuscript.

## References

[1] Alam, K. & Saxena, K. M. L. (1981). Positive dependence in multivariate distribution. Communications in Statistics–Theory and Methods, A 10, 1183–1196. https://doi.org/10.1080/03610928108828102

[2] Azevedo, C. & Oliveira, P. E. (2000). Kernel-type estimation of bivariate distribution function for associated random variables. New Trends in Probability and Statistics, 5, 17–25. https://doi.org/10.1515/9783110944655.17

[3] Chacon, J. E. & Rodriguez-Casal, A. (2010). A note on the universal consistency of the kernel distribution function estimator. Statistics and Probability Letters, 80, 1414–1419. https://doi.org/10.1016/j.spl.2010.05.007

[4] Donsker, M. D. (1951). An invariance principle for certain probability limit theorems. Memoirs of American Mathematical Society, 6, 1–12.

[5] Dubhashi, D., Priebe, V. & Ranjan, D. (1996). Negative dependence through the FKG inequality. Basic Research in Computer Science, RS-96-27, 1–15. https://doi.org/10.7146/brics.v3i27.20008

[7] Geenens, G. & Wang, C. (2018). Local-likelihood transformation kernel density estimation for positive random variables. Journal of Computational and Graphical Statistics, 27(4), 822–835. https://doi.org/10.1080/10618600.2018.1424636

[6] Ghosh, M. (1981). Multivariate negative dependence. Communications in Statistics–Theory and Methods, 10(4), 307–337. https://doi.org/10.1080/03610928108828041

[8] Hart, J. D. & Canette, I. (2011). Nonparametric estimation of distributions in random effects models. Journal of Computational and Graphical Statistic, 20(2), 461–478. https://doi.org/10.1198/jcgs.2011.09121

[9] Henriques, C. & Oliveira, P. E. (2003). Estimation of a two dimensional distribution function under association. Journal of Statistical Planning and Inference, 113(1), 137–150. https://doi.org/10.1016/S0378-3758(01)00296-8

[10] Henriques, C. & Oliveira, P. E. (2008). Strong convergence rates for the estimation of a covariance operator for associated samples. Statistical Inference for Stochastic Processes, 11, 77—91. https://doi.org/10.1007/s11203-006-9007-3

[11] Jabbari, H. (2009). Almost sure convergence of kernel bivariate distribution function estimator under negative association. Journal of Statistical Research of Iran, 6, 243–255. https://sid.ir/paper/129949/en

[12] Jabbari, H. (2013). On almost sure convergence for weighted sums of pairwise negatively quadrant dependent random variables. Statistical Papers, 54(3), 765–772. https://doi.org/10.1007/s00362-012-0460-3

[13] Jabbari, H. & Azarnoosh, H. A. (2006). Almost sure convergence rates for the estimation of a covariance operator for negatively associated samples. Journal of Iranian Statistical Society, 5(1-2), 53–67. https://sid.ir/paper/567630/en

[14] Jabbari Nooghabi, H. & Azarnoosh, H. A. (2009). Exponential inequality for negatively associated random variables. Statistical Papers, 50(2), 419–428. https://doi.org/10.1007/s00362-007-0081-4

[15] Joag-Dev, K. & Proschan, F. (1983). Negative association of random variables, with application. Annals of Statistics, 11(1), 286–295. https://doi.org/10.1214/aos/1176346079

[16] Koekemoer G. & Swanepoel Jan W. H. (2008). Transformation kernel density estimation with applications. Journal of Computational and Graphical Statistics, 17(3), 750–769. https://doi.org/10.1198/106186008X318585

[17] Matula, P. (1996). Probability and moment bounds for sums of negatively associated random variables. Theory of Probability and Mathematical Statistics, 55, 130–135. https://probability.knu.ua/tims/?year=1996& volume=55

[18] Miao, Yu., Zhao, F., Wang, K. & Chen, Y. (2013). Asymptotic normality and strong consistency of LS estimators in the EV regression model with NA errors. Statistical Papers, 54(1), 193–206. https://doi.org/10.1007/s00362-011-0418-x

[19] Newman, C. M. (1980). Normal fluctuations and the FKG inequalities. Communications in Mathematical Physics, 74, 119–128. https://doi.org/10.1007/BF01197754

[20] Newman, C. M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. Lecture Notes-Monograph Series, 5, 127–140. https://doi.org/10.1214/lnms/1215465639

[21] Sadikova, S. M. (1966). Two-dimensional analogies of an inequality of Esseen with applications to the central limit theorem. Theory of Probability and Its Applications, 11(3), 325–335. https://doi.org/10.1137/1111035

[22] Shao, Q. M. (2000). A comparison theorem on maximal inequality between negatively associated and independent random variables. Journal of Theoretical Probability, 13, 343–356. https://doi.org/10.1023/A:1007849609234

[23] Su, C. & Chi, X. (1988). Some results on CLT for nonstationary NA sequences. Acta Mathematicae Applicatae Sinica, 21(1), 9–21. https://applmath.cjoe.ac.cn/jweb_yysxxb_en

[24] Su, C., Zhao, L. C. & Wang, Y. B. (1997). Moment inequalities and weak convergence for negatively associated sequences. Science in China (series A), 40, 172–182. https://doi.org/10.1007/BF02874436

[25] Sung, S. H. (2011). On the strong convergence for weighted sums of random variables. Statistical Papers, 52(2), 447–454. https://doi.org/10.1007/s00362-009-0241-9

[26] Yu, M., Zhao, F., Wang, K. & Chen, Y. (2013). Asymptotic normality and strong consistency of LS estimators in the EV regression model with NA errors. Statistical Papers, 54(1), 193–206. https://doi.org/10.1007/s00362-011-0418-x

[27] Yuan, M., Su, C. & Hu, T. (2003). A central limit theorem for random fields of negatively associated process. Journal of Theoretical Probability, 16, 309–323. https://doi.org/10.1023/A:1023538824937

[28] Zarei, H. & Jabbari, H. (2011). Complete convergence of weighted sums under negative dependence. Statistical Papers, 52(2), 413–418. https://doi.org/10.1007/s00362-009-0238-4

[29] Zhou, Y. (1996). A note on the TJW product-limit estimator for truncated and censored data. Statistics and Probability Letters, 26(4), 381–387. https://doi.org/10.1016/0167-7152(95)00035-6

Hadi Jabbari
Orcid number: 0000-0002-7028-9052
Department of Statistics, Ordered Data, Reliability and Dependency Center of Excellence
Ferdowsi University of Mashhad
Mashhad, Iran
*Email address*: jabbarinh@um.ac.ir