# On almost sure convergence rates for the kernel estimator of a covariance operator under negative association

H. Jabbari[1] and M. Erfaniyan

*Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran*

**Abstract**

Let $\{X_n,\ n \geq 1\}$ be a strictly stationary sequence of negatively associated random variables, with common continuous and bounded distribution function $F$. We consider the estimation of the two-dimensional distribution function of $(X_1, X_{k+1})$ based on kernel type estimators as well as the estimation of the covariance function of the limit empirical process induced by the sequence $\{X_n,\ n \geq 1\}$ where $k \in \mathbb{N}_0$. Then, we derive uniform strong convergence rates for the kernel estimator of two-dimensional distribution function of $(X_1, X_{k+1})$ which were not found already and do not need any conditions on the covariance structure of the variables. Furthermore assuming a convenient decrease rate of the covariances $Cov(X_1, X_{n+1}),\ n \geq 1$, we prove uniform strong convergence rate for covariance function of the limit empirical process based on kernel type estimators. Finally, we use a simulation study to compare the estimators of distribution function of $(X_1, X_{k+1})$.

**Key Words**: Almost sure convergence rate, Bivariate distribution function, Empirical process, Kernel estimation.

# 1 Introduction, definitions and assumption

Estimation of distribution functions of random pairs (two-dimensional distribution functions) has been always a subject of interest of many statisticians. The case of independent underlying random variables was studied by [3]. The case of nonindependent random variables had been studied, too (see for example [1], [2], [6], [7] and [8]).

One of the most applicable concept of negative dependence in multivariate statistical analysis and reliability theory is negative association. 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 $\{1, 2, ..., n\}$,

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

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. We refer to [1], [8], [9], [10], [11], [12], [14], [15], [18], [19], [20], [21], [22], [23], [24] and [25] for knowing some of the most important studies have been performed on different aspects of NA random variables.

The mentioned comments above motivated the interest on 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

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

The asymptotic behavior of this estimator was studied by [6], [7] and [10]. For dependent sequences, under certain conditions (see [16], Theorem 17 and the first remark of p. 137), the

[1]E-mail: Jabbarinh@um.ac.ir

limit of the uniform empirical process still is a centered Gaussian process, but the covariance function changes to

$$\Gamma_k(r,s) \;=\; \varphi_k(r,s) + \sum_{k=1}^{\infty}\varphi_k(r,s) + \sum_{k=1}^{\infty}\varphi_k(s,r), \tag{2}$$

where $\varphi_k(r,s) = F_k(r,s) - F(r)F(s)$. [6], [7] and [10] drove 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 assuming a convenient decrease rate of the covariance. [2] and [8] considered the kernel estimator of $F_k$, defined by

$$\hat{F}_k(r,s) \;=\; \frac{1}{n-k}\sum_{i=1}^{n-k} U(\frac{r-X_i}{h_n}, \frac{s-X_{k+i}}{h_n}). \tag{3}$$

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 (3), we define the kernel estimator of $\varphi_k(r,s)$ and $\Gamma(r,s)$ as

$$\hat{\varphi}_k(r,s) = \hat{F}_k(r,s) - \hat{F}(r)\hat{F}(s), \;\; \hat{\Gamma}(r,s) = \hat{\varphi}_k(r,s) + \sum_{k=1}^{n}(\hat{\varphi}_k(r,s) + \hat{\varphi}_k(s,r)) \tag{4}$$

and derive a uniform convergence rate of order $h_{n-k}^2 n^{-\gamma}$ for the above estimators, where

$$\hat{F}(r) \;=\; \frac{1}{n}\sum_{i=1}^{n} U(\frac{r-X_i}{h_n})$$

and $0 < \gamma < 1/2$. For this 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 [10] and [8]. It is noted that the proofs are similar to those of [10]

In all sections of this paper suppose that $C$ is a positive constant not depending on $n$. Also, we use the following general assumption throughout the article:
**(A)**. $\{X_n,\ n \geq 1\}$ is a NA and strictly stationary sequence of random variables having bounded density function and

$$|U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) - EU(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n})| \;\leq\; Ch_n^2, \qquad a.s. \tag{5}$$

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

**Remark 1.1** *It can be easily checked that (5) holds for any NA sequence of random variables mentioned in* $(A)$, *because*

$$U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) \;=\; \int_{-\infty}^{Y_2}\int_{-\infty}^{Y_1} u(t_1,t_2)dt_1dt_2, \qquad a.s.$$

*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 letting* $z_1 = r - h_nt_1$ *and* $z_2 = s - h_nt_2$ *in the above integral, we have*

$$\begin{aligned} U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) \;&=\; \int_{-\infty}^{X_{i+k}}\int_{-\infty}^{X_i} u(\frac{r-z_1}{h_n}, \frac{s-z_2}{h_n})h_n^2 dz_1dz_2 \\ &\leq\; h_n^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(\frac{r-z_1}{h_n}, \frac{s-z_2}{h_n})dz_1dz_2. \qquad a.s. \end{aligned}$$

*By further replacements $w_1 = \frac{r-z_1}{h_n}$ and $w_2 = \frac{s-z_2}{h_n}$, we obtain*

$$U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) \quad = \quad O(h_n^2). \qquad\qquad a.s. \tag{6}$$

*On the other hand for the expected value of $U$, we can write*

$$EU(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) \quad = \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n})dU(x_i, x_{i+k}).$$

*By replacing $v_1 = \frac{r-X_i}{h_n}$ and $v_2 = \frac{s-X_{i+k}}{h_n}$, the above integral is equal to*

$$\begin{aligned} EU(\frac{r-X_i}{h_n}, \frac{s-X_{i+k}}{h_n}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(v_1, v_2)u(r - v_1 h_n, s - v_2 h_n)h_n^2 dv_1 dv_2 \\ &\leq h_n^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(r - v_1 h_n, s - v_2 h_n) dv_1 dv_2 \\ &= O(1). \end{aligned} \tag{7}$$

*The inequality of (7) holds, since $0 \leq U(v_1, v_2) \leq 1, \forall v_1, v_2 \in I\!R$ and the last equality holds after some more replacements. Finally, (5) satisfies by considering (6) and (7) together.*

In Section 2, we will present some auxiliary results needed to establish the above mentioned 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 a simulation study and then conclude the results.

# 2 Auxiliary results

In this section, we used the following moment inequality for NA random variables and proved an important inequality that are needed for proving our convergence rates.

**Lemma 2.1** *([13] and [20]) Let $(X_1, X_2, ..., X_n)$ be an NA random vector with $EX_j = 0$ and $E|X_j|^p < \infty$ for some $p \geq 2$ and all $j = 1, ..., n$. Then, there exists a constant $C = C(p) > 0$, such that*

$$E|\sum_{j=1}^{n} X_j|^p \quad \leq \quad C[\sum_{j=1}^{n} E|X_j|^p + (\sum_{j=1}^{n} EX_j^2)^{p/2}]. \qquad \square \tag{8}$$

**Lemma 2.2** *Let $k \in I\!N_0$ be fixed and $\varepsilon_n$ be a sequence of positive numbers. Suppose $(A)$ is satisfied. Then, there exists a constant $C$ such that, for $r, s \in I\!R$ and $p > 2$,*

$$P(|\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) \quad \leq \quad \frac{Ch_{n-k}^{2p}}{\varepsilon_n^p (n-k)^{p/2}}. \tag{9}$$

**Proof**. For each $n \in I\!N$, $1 \leq i \leq n$ and fixed $r, s \in I\!R$ define

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

and also

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

So, we have

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

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

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

Since (A) is hold, it is clear that $W_{k,n}$ are decreasing functions of the variables $X_n$. So according to the properties of NA random variables (see for more information [12]), $\{W_{k,n},\ n \geq 1\}$ 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

$$\begin{aligned}E|\sum_{i=1}^{n} W_{k,i}|^p &\leq C[\sum_{i=1}^{n} E|W_{k,i}|^p + (\sum_{i=1}^{n} EW_{k,i}^2)^{p/2}] \\ &\leq Cn^{p/2}h_n^{2p}. \qquad (10)\end{aligned}$$

Now for fixed $r, s \in I\!R$, we can write

$$\begin{aligned}P(|\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) &\leq P(|\hat{F}_k(r,s) - E(\hat{F}_k(r,s))| > \frac{\varepsilon_n}{2}) \\ &+ P(|F_k(r,s) - E(\hat{F}_k(r,s))| > \frac{\varepsilon_n}{2}). \qquad (11)\end{aligned}$$

Since $0 < F_k(r,s), \hat{F}_k(r,s) < 1$ for fixed $k \in I\!N_0$ and $r, s \in I\!R$, we conclude $P(|F_k(r,s) - E(\hat{F}_k(r,s))| > \frac{\varepsilon_n}{2}) \to 0$ as $n \to +\infty$. Now regarding this, using the Markov inequality and from (10) and (11) we find, for all $n > k$,

$$\begin{aligned}P(|\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) &\leq \frac{2^p}{\varepsilon_n^p(n-k)^p} E|\sum_{i=1}^{n-k} W_{k,i}|^p \\ &\leq \frac{Ch_{n-k}^{2p}}{\varepsilon_n^p(n-k)^{p/2}}. \qquad \square \qquad (12)\end{aligned}$$

To prove the next results, we should define the following notations as introduced in [10]. Let $t_n$ be a sequence of positive integers such that $t_n \to +\infty$. For each $n \in I\!N$ and each $i = 1, ..., t_n$, put $x_{n,i} = Q(i/t_n)$, where $Q$ is the quantile function of $F$. Then for $n \in I\!N$ and $k \in I\!N_0$, define

$$D_{n,k} = \sup_{r,s \in I\!R} |\hat{F}_k(r,s) - F_k(r,s)|,$$

and

$$D_{n,k}^* = \max_{i,j=1,\ldots,t_n} |\hat{F}_k(x_{n,i}, x_{n,j}) - F_k(x_{n,i}, x_{n,j})|.$$

Furthermore, we will need the following result as in Theorem 2 of [6] and Lemma 2.3 of [10].

**Lemma 2.3** *If the sequence $\{X_n,\ n \geq 1\}$ satisfies (A), then, for each $n \in I\!N$ and each $k \in I\!N_0$,*

$$D_{n,k} \leq D_{n,k}^* + \frac{2}{t_n} \qquad a.s.\ . \qquad \square \qquad (13)$$

**Lemma 2.4** *Let $\varepsilon_n$ and $t_n$ be two sequences of positive numbers such that $t_n \to +\infty$ and $\varepsilon_n t_n \to +\infty$, $p > 2$ and $k \in \mathbb{N}_0$ be fixed. Suppose (A) holds. Then, for any large enough $n$,*

$$P(\sup_{r,s\in\mathbb{R}} |\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) \leq \frac{Ct_n^2}{\varepsilon_n^p (n-k)^{p/2}} h_{n-k}^{2p}. \tag{14}$$

**Proof**. Following the same steps in Lemma 2.4 of [10] and applying Lemma 2.2 and Lemma 2.3 the result is concluded. □

# 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 $k \in \mathbb{N}_0$ be fixed and suppose (A) holds. Then under the conditions of Lemma 2.4 and for every $0 < \delta < p-1$, we have*

$$\sup_{r,s\in\mathbb{R}} |\hat{F}_k(r,s) - F_k(r,s)| = O(h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{p-2-2\delta}{2(p+2)}}) \qquad a.s. . \tag{15}$$

**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 2.4 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

$$P(\sup_{r,s\in\mathbb{R}} |\hat{F}_k(r,s) - F_k(r,s)| > \varepsilon_n) \leq \frac{C}{\varepsilon_n^{p+2} h_{n-k}^{2-2p} (n-k)^{p/2}} \leq Cn^{-(1+\delta)}. \tag{16}$$

The proof is complete using the Borel-Cantelli Lemma, because for all $\delta > 0$, the sequence on the right-hand side above being summable. □

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

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

$$\sup_{r,s\in\mathbb{R}} |\hat{F}_k(r,s) - F_k(r,s)| = O(h_{n-k}^2 n^{-\gamma}) \qquad a.s. . \tag{17}$$

**Proof**. Using Lemma 3.1 and along the lines of Theorem 3.1 in [10], we get the desired result. □

**Remark 3.1** *Note that Theorem 4 of [8] holds true for $\hat{F}_k$ defined in (3) under some regularity assumptions. So for all $x, y \in \mathbb{R}$, we have*

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

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

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

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

If $k = 0$ and $s = r$ the estimator $\hat{F}_k(r,s)$ becomes to the one-dimentional kernel distribution function $\hat{F}(r)$. The results of Theorem 3.1 hold true for $\hat{F}$. So, we can write

$$\sup_{r\in \mathbb{R}} |\hat{F}(r) - F(r)| \;=\; O(h_n^2 n^{-\gamma}) \qquad a.s.\,. \tag{18}$$

**Remark 3.2** *From the results of Theorem 3.1, 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 [10] (i.e. $n^{-\gamma}$). So, the kernel estimator of two-dimensional and one-dimentional distribution function $F_k$ and $F$ is better than empirical one, respectively.*

Now, we can obtain the convergence rate of the kernel estimator of $\varphi_k$.

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

$$\sup_{r,s\in \mathbb{R}} |\hat{\varphi}_k(r,s) - \varphi_k(r,s)| \;=\; O(h_{n-k}^2 n^{-\gamma}) \qquad a.s.\,. \tag{19}$$

**Proof**. The proof is similar to that of Theorem 3.2 in [10] and then we omit it.

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

As [10], we will introduce uniform strong convergence rates for the kernel estimators of the sum $\sum_{k=1}^{\infty}\varphi_k(r,s)$ and the covariance function $\Gamma(r,s)$.
Regarding that the covariance structure of a sequence of NA random variables highly determines its approximate independence (see [16]), it is common to have assumptions on the covariance structure of the random variables. For this, we use the same definition of [10] as

$$v(n) \;=\; \sum_{j=n+1}^{\infty} |Cov(X_1, X_j)|^{1/3}. \tag{20}$$

In the following lemma, 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 $\Gamma$.

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

$$v(a_n) \;\leq\; C h_{n-k}^{\frac{4\theta(p-1)}{(p-2)(p+3)}} a_n^{-\theta} \tag{21}$$

*for all $n \geq 1$, we have*

$$\sup_{r,s\in \mathbb{R}} |\sum_{k=1}^{a_n}\hat{\varphi}_k(r,s) - \sum_{k=1}^{\infty}\varphi_k(r,s)| \;=\; O(h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}) \qquad a.s.\,. \tag{22}$$

**Proof**. The idea is essentially the same as the proof of Lemma 4.1 of [10]. So, we repeat their proof using our required notations and definitions.
Take $\varepsilon_n = h_{n-k}^{\frac{2p-2}{p+2}} n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}$ for each $0 < \delta < \frac{p-2}{2}$ and $t_n = \frac{a_n}{\varepsilon_n h_{n-k}}$. Now, we can write

$$P(\sup_{r,s\in \mathbb{R}} |\sum_{k=1}^{a_n}(\hat{F}_k(r,s) - F_k(r,s))| > \varepsilon_n) \;\leq\; \sum_{k=1}^{a_n} P(\sup_{r,s\in \mathbb{R}} |\hat{F}_k(r,s) - F_k(r,s)| > \frac{\varepsilon_n}{a_n}). \tag{23}$$

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 $\varepsilon_n \to 0$, $a_n \to +\infty$, $t_n \to +\infty$, $\frac{\varepsilon_n}{a_n} t_n \to +\infty$ and $\frac{a_n}{n} \to 0$ as $n \to +\infty$.

Using $\frac{\varepsilon_n}{a_n}$ in place of $\varepsilon_n$ in Lemma 2.4, we obtain for all $n$ large enough,

$$\begin{aligned} P(\sup_{r,s\in I\!\!R}|\sum_{k=1}^{a_n}(\hat{F}_k(r,s)-F_k(r,s))|>\varepsilon_n) &\leq \sum_{k=1}^{a_n}\frac{Ct_n^2a_n^p}{\varepsilon_n^p(n-k)^{p/2}}h_{n-k}^{2p} \\ &\leq \frac{Ct_n^2a_n^{p+1}}{\varepsilon_n^p(n-a_n)^{p/2}}h_{n-k}^{2p} \\ &= \frac{Ca_n^{p+3}}{\varepsilon_n^{p+2}(n-a_n)^{p/2}}h_{n-k}^{2p-2}. \end{aligned} \quad (24)$$

By elementary 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 (24) leads to summable upper bound as $\frac{a_n}{n}\to 0$. So, we have by Borel-Cantelli Lemma

$$\sup_{r,s\in I\!\!R}|\sum_{k=1}^{a_n}(\hat{F}_k(r,s)-F_k(r,s))| = O(h_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}) \quad a.s.\,. \quad (25)$$

Now, as [10], we can write

$$\begin{aligned} \sup_{r,s\in I\!\!R}|\sum_{k=1}^{a_n}\hat{\varphi}_k(r,s)-\sum_{k=1}^{\infty}\varphi_k(r,s)| &\leq \sup_{r,s\in I\!\!R}|\sum_{k=1}^{a_n}(\hat{F}_k(r,s)-F_k(r,s))| \\ &+ 2a_n\sup_{r\in I\!\!R}|\hat{F}(r)-F(r)| \\ &+ \sup_{r,s\in I\!\!R}|\sum_{k=a_n+1}^{\infty}\varphi_k(r,s)|. \end{aligned} \quad (26)$$

For the first term on the right-hand side of (26), we use (25). Since $\frac{p+3}{p+2}>1$ by using Lemma 3.1 for the second term, we have

$$\begin{aligned} a_n\sup_{r\in I\!\!R}|\hat{F}(r)-F(r)| &= O(a_nh_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{p-2-2\delta}{2p+4}}) \\ &= O(h_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}) \quad a.s.\,. \end{aligned} \quad (27)$$

For the third term on the right-hand side of (26), we use Corollary of Theorem 1 in [17] and relation (21) in [15] as those applied in [10]. So by (21) for $\theta=\frac{(p-2)(p+3)}{2p+4}>0$ and $a_n = n^{\frac{p-2-2\delta}{p^2+3p}}$, we obtain

$$\begin{aligned} \sup_{r,s\in I\!\!R}|\sum_{k=a_n+1}^{\infty}\varphi_k(r,s)| &\leq C\sum_{k=a_n+1}^{\infty}|Cov^{1/3}(X_1,X_{k+1})| \\ &= Cv(a_n)\leq Ch_{n-k}^{\frac{4(p-1)(p-2)(p+3)}{2(p+2)(p-2)(p+3)}}a_n^{-\frac{(p-2)(p+3)}{2(p+2)}} \\ &= Ch_{n-k}^{\frac{2p-2}{p+2}}n^{-\frac{(p-2)(p-2-2\delta)}{2p(p+2)}}. \end{aligned} \quad (28)$$

Hence the proof is completed. □

We now summarize the above result in the following theorem.

**Theorem 4.1** *Under the assumptions of Lemma 4.1 and condition (21) for all $n\geq 1$, $\theta>0$ and $0<\gamma<1/2$, we have*

$$\sup_{r,s\in I\!\!R}|\sum_{k=1}^{a_n}\hat{\varphi}_k(r,s)-\sum_{k=1}^{\infty}\varphi_k(r,s)| = O(h_{n-k}^2n^{-\gamma}) \quad a.s.\,. \quad (29)$$

**Proof**. As in proof of Theorem 4.1 of [10], we apply the lines of proof of Theorem 3.1 and use Lemma 4.1 instead of Lemma 3.1. So, for $\delta>0$ and $p>2$ we have $\frac{(p-2)(p-2-2\delta)}{2p(p+2)}>\gamma$ and then the proof is concluded. □

Now, applying the lines of proof of Theorem 4.2 in [10] and using Theorems 3.1 and 4.1, we can state the following theorem which summarizes the results for $\hat{\Gamma}$.

**Theorem 4.2** *Suppose (A) holds. Under condition (21) for all $n \geq 1$, $\theta > 0$, $p > 2$ and $0 < \gamma < 1/2$, we have*

$$\sup_{r,s \in I\!R} |\hat{\Gamma}(r,s) - \Gamma(r,s)| \quad = \quad O(h_{n-k}^2 n^{-\gamma}) \qquad a.s. \ . \quad \square \tag{30}$$

**Remark 4.1** *As stated in Remark 3.2, our convergence rate $h_{n-k}^2 n^{-\gamma}$ for every $0 < \gamma < 1/2$ and $h_n$ in Theorem 4.2 is very faster than those obtained later by [10] (i.e. $n^{-\gamma}$). So, the kernel estimator of $\Gamma$ is better than empirical one.*

# 5 Simulation study

In this section, we intend to compare the behavior of our estimator with those of [10] via a simulation study. As noted in [4], [5] and [12] a number of well known multivariate distributions such as multivariate normal distribution with negative correlations possess the NA property. So for generating the NA random variables, suppose that $X_1, \ldots, X_n$ have multivariate normal joint distribution with zero mean vector and the following covariance matrix

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

where $\rho > 0$. For $n = 20, 100$, we generate one sample from $n$-dimensional multivariate normal distribution with zero mean vector and covariance matrix $\Sigma$ 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 (3) using $h_n = n^{-1}$ and $h_n = \log^{-1}(n)$ and $U(.,.)$ as bivariate normal distribution with zero mean vector and covariance matrix

$$\frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}. \tag{31}$$

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

$$\begin{aligned} MSD_1 &= \frac{1}{N} \sum_{r,s} (\hat{F}_k(r,s) - F_k(r,s))^2 \\ MSD_2 &= \frac{1}{N} \sum_{r,s} (\tilde{F}_k(r,s) - F_k(r,s))^2 \end{aligned} \tag{32}$$

where $N$ is the product of all numbers $r$ and $s$. The results are also reported in Figures 1-3.

Figure 1 shows 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 rates $h_n = \log^{-1}(n)$ 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.

Figure 2 shows 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.
**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 $F_1(r, s)$.

Figure 3 shows that 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.

# References


[1] Alam, K. and Saxena, K.M.L. (1981). Positive dependence in multivariate distribution. *Comm. Stat-Theor. Meth.*, A 10, 1183–1196.

[2] Azevedo, C. and Oliveira, P.E. (2000). Kernel-type estimation of bivariate distribution function for associated random variables. *New Trends in Probability and Statistics*, 17–25.

[3] Donsker, M.D. (1951). An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6, 1–12.

[4] Dubhashi, D., Priebe, V. and Ranjan, D. (1996). Negative dependence through the FKG inequality. *Basic Research in Computer Science*, RS-96-27, 1–15.

[5] Ebrahimi, N. and Ghosh, M. (1981). Multivariate negative dependence. *Comm. Stat-Theor. Meth.*, 4, 307–337.

[6] Henriques, C. and Oliveira, P.E. (2003). Estimation of a two dimensional distribution function under association. *J. Statist. Planning Inf.*, 113, 137–150.

[7] Henriques, C. and Oliveira, P.E. (2005). Strong convergence rates for the estimation of a covariance operator for associated samples. Preprint, Pre-Publicacoes do Departamento de Matematica da Universidade de Coimbra 05–05.

[8] Jabbari, H. (2009). Almost sure convergence of kernel bivariate distribution function estimator under negative association. *Journal of Statistical Research of Iran*, 6, 243–255.

[9] Jabbari, H. (2013). On almost sure convergence for weighted sums of pairwise negatively quadrant dependent random variables. *Stat. Papers*, 54(3), 765–772.

[10] Jabbari, H. and 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.

[11] Jabbari Nooghabi, H. and Azarnoosh, H.A. (2009). Exponential inequality for negatively associated random variables. *Stat. Papers*, 50(2), 419–428.

[12] Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with application. *Ann. Stat.*, 11, 286–295.

[13] Matula, P. (1997). Probability and moment bounds for sums of negatively associated random variables. *Theory Probab. Math. Stat.*, 55, 135–141.

[14] Miao, Yu. Zhao, F., Wang, K. and Chen, Y. (2013). Asymptotic normality and strong consistency of LS estimators in the EV regression model with NA errors. *Stat. Papers*, 54(1), 193–206.

[15] Newman, C.M. (1980). Normal fluctuations and the FKG inequalities. *Comm. Math. Phys.*, 74, 119–128.

[16] Newman, C. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables, in: Inequalities in Statistics and Probability, Y.L. Tong, ed, IMS Lecture Notes- Monograph Series, 5, 127–140.

[17] Sadikova, S.M. (1966). Two-dimensional analogies of an inequality of Esseen with applications to the central limit theorem. *Theory Probab. Appl.*, 11, 325–335.

[18] Shao, Q.M. (2000). A comparison theorem on maximal inequality between negatively associated and independent random variables. *J. Theor. Probab.*, 13, 343–356.

[19] Su, C. and Chi, X. (1998). Some result on CLT for nonstationary NA sequences. *Acta. Math. Appl. Sinica*, 21, 9–21.

[20] Su, C., Zhao, L.C. and Wang, Y.B. (1997). Moment inequalities and weak convergence for negatively associated sequences. *Science in China (series A)*, 40, 172–182.

[21] Sung, S.H. (2011). On the strong convergence for weighted sums of random variables. *Stat. Papers*, 52(2), 447–454.

[22] Yu, M., Zhao, F., Wang, K. and Chen, Y. (2013). Asymptotic normality and strong consistency of LS estimators in the EV regression model with NA errors. *Stat. Papers*, 54(1), 193–206.

[23] Yuan, M., Su, C. and Hu, T. (2003). A central limit theorem for random fields of negatively associated process. *J. Theor. Probab.*, 16, 309–323.

[24] Zarei, H. and Jabbari, H. (2011). Complete convergence of weighted sums under negative dependence. *Stat. Papers*, 52(2), 413–418.

[25] Zhou, Y. (1996). A note on the TJW product-limit estimator for truncated and censored data. *Statist. Probab. Lett.*, 26, 381–387.