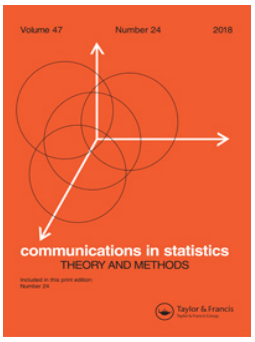

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Exponential convergence rates for the kernel bivariate distribution function estimator under NSD assumption with application to hydrology data

A. Kheyri , M. Amini , H. Jabbari , A. Bozorgnia & A. Volodin

To cite this article: A. Kheyri , M. Amini , H. Jabbari , A. Bozorgnia & A. Volodin (2020): Exponential convergence rates for the kernel bivariate distribution function estimator under NSD assumption with application to hydrology data, Communications in Statistics - Theory and Methods

To link to this article: https://doi.org/10.1080/03610926.2020.1808900

Published online: 01 Sep 2020.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2020.1808900

Exponential convergence rates for the kernel bivariate distribution function estimator under NSD assumption with application to hydrology data

A. Kheyri[a], M. Amini[a], H. Jabbari[a], A. Bozorgnia[b], and A. Volodin[c]

[a]Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran; [b]Department of Statistics, Khayyam University, Mashhad, Iran; [c]Department of Mathematics and Statistics, University of Regina, Regina, Canada

ABSTRACT

In this paper, we study the asymptotic behavior of the kernel bivariate distribution function estimator for negatively superadditive dependent. The exponential convergence rates for the kernel estimator are investigated. Under certain regularity conditions, the optimal bandwidth rate is determined with respect to mean squared error criteria. A simulation study is used to justify the behavior of the kernel and histogram estimators. As an application, a real data set in hydrology is considered and the kernel bivariate distribution function estimator of the data is investigated.

ARTICLE HISTORY
Received 28 December 2019
Accepted 6 August 2020

KEYWORDS
Bivariate distribution function; kernel estimation; negative superadditive dependence

MSC
Primary 62G07;
Secondary 62G20

1. Introduction

The estimation of bivariate distribution functions has been a subject of interest in a large volume of statistical literature. There has been extensive work on the statistical estimation of the bivariate distribution function using the kernel method. Several properties of the kernel type estimator have been investigated. For example, Donsker (1951) examined the case of independent random variables, whereas, Azevedo and Oliveira (2000), Henriques and Oliveira (2003, 2008), and Jabbari (2009) examined dependent random variables.

One of the most applicable dependence concepts is that of negative superadditive dependence (NSD), which was introduced by Hu (2000). The definition of NSD random variables is expressed on the basis of the superadditive functions. A function $\phi : R^n \to R$ is called superadditive if

$$\phi(\mathbf{x}\vee\mathbf{y}) + \phi(\mathbf{x}\wedge\mathbf{y}) \geq \phi(\mathbf{x}) + \phi(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in R^n$, where $\vee$ and $\wedge$ stand for componentwise maximum and minimum, respectively. Consequently, the NSD concept is expressed as follows. A random vector $(X_1, \ldots, X_n)$ is said to be NSD if

CONTACT Mohammad Amini m-amini@um.ac.ir Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran.

$$E\phi(X_1,...,X_n) \le E\phi(X_1^*,...,X_n^*), \tag{1.1}$$

where $X_1^*,...,X_n^*$ are independent such that X_i^* and X_i have the same distribution for each i, and $\phi(\cdot)$ is a superadditive function such that the expectations above exist. If $\phi(\cdot)$ has continuous second partial derivatives, then the superadditivity of $\phi(\cdot)$ is equivalent to $\partial^2\phi/\partial x_i\partial x_j \ge 0, 1 \le i \ne j \le n$. Also, a sequence $\{X_n, n \ge 1\}$ of random variables is NSD if every finite subfamily is NSD.

In this paper, we consider some samples that satisfy the notion of NSD. Christofides and Vaggelatou (2004) showed that the family of NSD sequences contains negatively associated (NA) random variables as a special case. Therefore, the probability inequalities obtained based on the NSD assumption is more general. Numerous limit theorems for NSD random variables have been studied; some recent works are Chen et al. (2020), Cong, Tran, and Le (2020), and Kheyri et al. (2019a, 2019b).

The most common estimator of $F_k(x,y) = P(X_1 \le x, X_{k+1} \le y)$ with k fixed, constructed on the basis of the first n random variables from the sequence, is the histogram estimator, $\tilde{F}_k(x,y)$ that is defined by

$$\tilde{F}_k(x,y) = \frac{1}{n-k}\sum_{i=1}^{n-k} I(X_i \le x)I(X_{k+i} \le y). \tag{1.2}$$

The asymptotic behavior of this estimator was studied for associated random variables by Henriques and Oliveira (2003, 2008), and Jabbari and Azarnoosh (2006). Here, we consider the kernel estimator that was used by Azevedo and Oliveira (2000) and Jabbari (2009), in the following

$$\hat{F}_k(x,y) = \frac{1}{n-k}\sum_{i=1}^{n-k} U\left(\frac{x-X_i}{h_n}, \frac{y-X_{k+i}}{h_n}\right), \tag{1.3}$$

where $U(\cdot,\cdot)$ is a given bivariate distribution function and $\{h_n, n \ge 0\}$ is a sequence of positive numbers converging to zero. They studied the uniform convergence of the kernel estimator for associated random variables and found the corresponding optimal bandwidth convergence rate. Here, we extend and improve their results to NSD random variables.

The remaining sections of the paper are organized as follows. The exponential inequality and uniform convergence rate of the kernel estimator under NSD are introduced in the next section, followed by the convergence rate of the kernel estimator. Some asymptotic properties and convergence rates of the mean square error (MSE) are studied in Section 3. Moreover, we illustrate the behavior of the kernel and histogram estimators with respect to their empirical mean square distances (MSDs) in Section 4. In Section 5, an application to rainfall depth data is considered: data that were discussed later by Kheyri et al. (2019a).

2. An exponential convergence rate

In this section, we prove an exponential convergence rate for the kernel estimator of $F_k(x,y)$ with k fixed. All results are derived under the basic assumption of NSD. The remaining assumptions that need to prove the main results are listed below.

A1(i) The sequence $\{X_n, n \geq 1\}$ is an identically distributed sequence of NSD random variables with the bounded density function $f(\cdot)$.
A1(ii) The bivariate distribution function $F_k(\cdot,\cdot)$ has bounded and continuous partial derivatives of first and second orders.
A1(iii) The distribution function of $(X_1, X_{k+1}, X_j, X_{k+j})$, denoted by $F_{kj}(\cdot,\cdot,\cdot,\cdot)$, has bounded and continuous partial derivatives of first and second orders.
A2 The sequence of bandwidth $\{h_n; n > 1\}$ is such that, as $n \to \infty$:

(i) $0 < h_n \to 0$ **(ii)** $nh_n^2 \to 0$

A3 The bivariate function $U(\cdot,\cdot)$ is a bivariate distribution function with density function $u(\cdot,\cdot)$. $U(\cdot,\cdot)$ is twice differentiable and

$$\int_{R^2} xu(x,y)dxdy = \int_{R^2} yu(x,y)dxdy = 0,$$
$$\int_{R^2} x^2u(x,y)dxdy < \infty, \int_{R^2} y^2u(x,y)dxdy < \infty.$$

A4 The sequence $\{X_n, n \geq 1\}$ is an identically distributed sequence of NSD random variables and there is a constant C for which

$$\left|U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right) - E\left[U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right)\right]\right| \leq Ch_n^2 \quad a.s. \tag{2.1}$$

Remark 1. Assumptions **A1** and **A2** are often applied to the asymptotic theory of kernel estimators in the literature. **A3** satisfies some common bivariate kernel functions, for example, the bivariate normal distribution or bivariate Farlie–Gumbel–Morgenstern distribution when the marginal density functions have zero mean and finite variance. Also, **A4** is reasonable because of the following statement:

$$U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right) - E\left[U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right)\right] = \int_{-\infty}^{\frac{x-X_i}{h_n}} \int_{-\infty}^{\frac{y-X_{i+k}}{h_n}} u(r,s)drds$$
$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U\left(\frac{x-w}{h_n}, \frac{y-v}{h_n}\right) dF_k(w,v). \qquad a.s.$$

By letting $r = x + h_nr^*, s = y + h_ns^*$ and $w = x - h_nw^*, v = y - h_nv^*$ and doing some calculations, we have

$$\left|U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right) - E\left[U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right)\right]\right| \leq h_n^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x+h_nr^*, y+h_ns^*)dr^*ds^*$$
$$+ h_n^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(w^*, v^*)dw^*dv^* = O(h_n^2). \qquad a.s.$$

Lemma 2.1. (Hoeffding, 1963). Let X be a random variable with $E(X) = \mu$. If there exist $a, b \in R$ such that $P(a \leq X \leq b) = 1$, then for every $\lambda > 0$,

$$E(e^{\lambda X}) \leq e^{\lambda\mu} \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Proposition 2.1. *If A1(i), A2(i), A3 and A4 hold, then for every $\varepsilon > 0$,*

$$P\left(\sup_{x,y\in R}\left|\hat{F}_k(x,y) - E\left[\hat{F}_k(x,y)\right]\right| > \varepsilon\right) \leq 2\exp\left(-\frac{(n-k)\varepsilon^2}{2C^2h_n^4}\right). \tag{2.2}$$

Proof. For k fixed and $i = 1,\dots,n$, let

$$T_{i,k,n}(x,y) = U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right) - E\left[U\left(\frac{x-X_i}{h_n}, \frac{y-X_{i+k}}{h_n}\right)\right],$$

then for every $\varepsilon > 0$,

$$\begin{aligned}
&P\left(\sup_{x,y\in R}\left|\hat{F}_k(x,y) - E\left[\hat{F}_k(x,y)\right]\right| > \varepsilon\right) = P\left(\sup_{x,y\in R}\left|\sum_{i=1}^{n-k} T_{i,k,n}(x,y)\right| > (n-k)\varepsilon\right)\\
&= P\left(\sup_{x,y\in R}\left[\left(\sum_{i=1}^{n-k} T_{i,k,n}(x,y)\right)^+ + \left(\sum_{i=1}^{n-k} T_{i,k,n}(x,y)\right)^-\right] > (n-k)\varepsilon\right)\\
&\leq P\left(\sup_{x,y\in R}\left(\sum_{i=1}^{n-k} T_{i,k,n}(x,y)\right)^+ > \frac{(n-k)\varepsilon}{2}\right) + P\left(\sup_{x,y\in R}\left(\sum_{i=1}^{n-k} T_{i,k,n}(x,y)\right)^- > \frac{(n-k)\varepsilon}{2}\right)\\
&\leq P\left(\sup_{x,y\in R}\sum_{i=1}^{n-k} T^+_{i,k,n}(x,y) > \frac{(n-k)\varepsilon}{2}\right) + P\left(\sup_{x,y\in R}\sum_{i=1}^{n-k} T^-_{i,k,n}(x,y) > \frac{(n-k)\varepsilon}{2}\right),
\end{aligned} \tag{2.3}$$

where $T^+_{i,k,n}(x,y) = \max(T_{i,k,n}(x,y), 0)$ and $T^-_{i,k,n}(x,y) = \max(-T_{i,k,n}(x,y), 0)$. So for all $t > 0$; using Markov's inequality, we can write

$$\begin{aligned}
&P\left(\sup_{x,y\in R}\sum_{i=1}^{n-k} T^+_{i,k,n}(x,y) > \frac{(n-k)\varepsilon}{2}\right) \leq P\left(\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y) > \frac{(n-k)\varepsilon}{2}\right)\\
&\leq e^{-\frac{(n-k)\varepsilon t}{2}} E\left(\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y)\right)\right).
\end{aligned} \tag{2.4}$$

If $\phi(x_1,\dots,x_n) = \exp\left(t\sum_{z=1}^{n-k} g(x_z, x_{z+k})\right)$, where $g(x_z, x_{z+k}) = \sup_{x,y\in R} T^+_{z,k,n}(x,y)$ for any $i \neq j$, $\partial^2\phi(x_1,\dots,x_n)/\partial x_i\partial x_j$ will be in one of the following forms:

1) $$t^2\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\frac{\partial g(x_j,x_{j+k})}{\partial x_j}\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y)\right),$$

2) $$t^2\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\frac{\partial g(x_{j-k},x_j)}{\partial x_j}\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y)\right),$$

3) $$t^2\frac{\partial g(x_{i-k},x_i)}{\partial x_i}\frac{\partial g(x_{j-k},x_j)}{\partial x_j}\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y)\right),$$

4) $$t^2\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\left(\frac{\partial g(x_{j-k},x_j)}{\partial x_j} + \frac{\partial g(x_j,x_{j+k})}{\partial x_j}\right)\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y)\right),$$

5) $$t^2\frac{\partial g(x_{j-k},x_j)}{\partial x_j}\left(\frac{\partial g(x_{i-k},x_i)}{\partial x_i} + \frac{\partial g(x_i,x_{i+k})}{\partial x_i}\right)\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R} T^+_{i,k,n}(x,y)\right),$$

$$6)\left[t\frac{\partial^2 g(x_i,x_{i+k})}{\partial x_i\partial x_{i+k}}+t^2\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\left(\frac{\partial g(x_i,x_{i+k})}{\partial x_{i+k}}+\frac{\partial g(x_{i+k},x_{i+2k})}{\partial x_{i+k}}\right)\right]\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R}T^{+}_{i,k,n}(x,y)\right),$$

$$7)\left[t\frac{\partial^2 g(x_i,x_{i+k})}{\partial x_i\partial x_{i+k}}+t^2\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\left(\frac{\partial g(x_{i-k},x_i)}{\partial x_i}+\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\right)\right]\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R}T^{+}_{i,k,n}(x,y)\right),$$

$$8)\ t^2\left(\frac{\partial g(x_{i-k},x_i)}{\partial x_i}+\frac{\partial g(x_i,x_{i+k})}{\partial x_i}\right)\left(\frac{\partial g(x_{j-k},x_j)}{\partial x_j}+\frac{\partial g(x_j,x_{j+k})}{\partial x_j}\right)\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R}T^{+}_{i,k,n}(x,y)\right).$$

It can be shown that the sign of each of the above statements is non negative. Therefore, $\phi(x_1,...,x_n)$ is a superadditive function; so

$$E\left(\exp\left(t\sum_{i=1}^{n-k}\sup_{x,y\in R}T^{+}_{i,k,n}(x,y)\right)\right)\le\prod_{i=1}^{n-k}E\left(\exp\left(t\sup_{x,y\in R}T^{+}_{i,k,n}(x,y)\right)\right). \tag{2.5}$$

Consequently by A4, (2.4), (2.5) and using Lemma 2.1, we have

$$\begin{aligned}P\left(\sup_{x,y\in R}\sum_{i=1}^{n-k}T^{+}_{i,k,n}(x,y)>\frac{(n-k)\varepsilon}{2}\right)&\le\exp\left(\frac{-(n-k)\varepsilon t}{2}\right)\prod_{i=1}^{n-k}E\left(\exp\left(t\sup_{x,y\in R}T^{+}_{i,k,n}(x,y)\right)\right)\\&\le\exp\left(-tE(\sup_{x,y\in R}T^{+}_{i,k,n}(x,y))+\frac{(n-k)t^2C^2h_n^4}{8}\right).\end{aligned}$$

Since $0<E(\sup_{x,y\in R}T^{+}_{i,k,n}(x,y))<\infty$, therefore

$$P\left(\sup_{x,y\in R}\sum_{i=1}^{n-k}T^{+}_{i,k,n}(x,y)>\frac{(n-k)\varepsilon}{2}\right)\le\exp\left(\frac{-(n-k)\varepsilon t}{2}+\frac{(n-k)t^2C^2h_n^4}{8}\right). \tag{2.6}$$

Now by minimizing the right-hand side of (2.6) with respect to t and substituting the optimal bound, we obtain that

$$P\left(\sup_{x,y\in R}\sum_{i=1}^{n-k}T^{+}_{i,k,n}(x,y)>\frac{(n-k)\varepsilon}{2}\right)\le\exp\left(-\frac{(n-k)\varepsilon^2}{2C^2h_n^4}\right). \tag{2.7}$$

Similarly, an optimal bound for the last term in (2.3) is achieved, and the proof is complete. □

In fact, we derived sufficient conditions to prove an exponential rate for the kernel-type estimator of the distribution function. To prove the convergence rate, we choose ε depending on n as

$$\varepsilon_n^2=\frac{\alpha h_n^4\log(n-k)}{n-k},$$

in order to obtain a convergence series in the right-hand side of (2.2), where $\alpha>0$ must be conveniently chosen (it depends on constants appearing in the inequality). So, the convergence rate of the kernel-type estimator is of the order $O(\sqrt{h_n^4\log(n-k)/(n-k)})$.

3. The mean square error

Here, we study the asymptotic properties and convergence rate of the MSE of the estimator. From this, we derive the optimal bandwidth rate of the order $n^{-1/3}$.

Lemma 3.1. *Let A1(i)(ii), A2(i) and A3 be satisfied. Then,*

$$\hat{F}_k(x,y) = F_k(x,y) + O(h_n^2), \tag{3.1}$$

$$\mathrm{Var}\left(U\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right)\right) = F_k(x,y) - F_k^2(x,y) + O(h_n), \tag{3.2}$$

and if A1(iii) also is hold, then we have for $j \neq 1$

$$\mathrm{Cov}\left(U\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), U\left(\frac{x-X_j}{h_n}, \frac{y-X_{k+j}}{h_n}\right)\right) = F_{kj}(x,y,x,y) - F_k^2(x,y) + O(h_n^2). \tag{3.3}$$

Proof. Using integration by parts and the Taylor expansion, the expectation of the kernel estimator can be derived as

$$\begin{aligned}
E(\hat{F}_k(x,y)) &= \int_{R^2} U\left(\frac{x-r}{h_n}, \frac{y-s}{h_n}\right) dF_k(r,s) \\
&= \int_{R^2} U\left(\frac{x-r}{h_n}, \frac{y-s}{h_n}\right) \frac{\partial^2 F_k(r,s)}{\partial r \partial s} drds \\
&= -\int_{R^2} \frac{\partial U\left(\frac{x-r}{h_n}, \frac{y-s}{h_n}\right)}{\partial r} \frac{\partial F_k(r,s)}{\partial s} dsdr \\
&= \int_{R^2} \frac{\partial^2 U\left(\frac{x-r}{h_n}, \frac{y-s}{h_n}\right)}{\partial r \partial s} F_k(r,s) dsdr \\
&= \int_{R^2} u(r,s) F_k(x - h_n r, y - h_n s) dsdr \\
&= F_k(x,y) + \frac{h_n^2}{2} F_{k(11)}(x,y) \int_{R^2} r^2 u(r,s) dsdr \\
&\qquad + \frac{h_n^2}{2} F_{k(22)}(x,y) \int_{R^2} s^2 u(r,s) dsdr + o(h_n^2) \\
&= F_k(x,y) + O(h_n^2),
\end{aligned}$$

where $F_{k(11)}(x,y)$ and $F_{k(22)}(x,y)$ are the second order partial derivatives with respect to the first and second components respectively at point (x, y).

Similarly, the variance of the kernel estimator can be written as

$$
\begin{aligned}
\operatorname{Var}\left(U\left(\frac{x-X_1}{h_n},\frac{y-X_{k+1}}{h_n}\right)\right) &= E\left(U\left(\tfrac{x-X_1}{h_n},\tfrac{y-X_{k+1}}{h_n}\right)\right)^2 - E^2(\hat{F}_k(x,y))\\
&= \int_{R^2} U^2\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)dF_k(r,s) - E^2(\hat{F}_k(x,y))\\
&= \int_{R^2} U^2\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)\frac{\partial^2 F_k(r,s)}{\partial r\partial s}drds - E^2(\hat{F}_k(x,y))\\
&= -\int_{R^2} \frac{\partial U^2\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)}{\partial r}\frac{\partial F_k(r,s)}{\partial s}drds - E^2(\hat{F}_k(x,y))\\
&= \int_{R^2} \frac{\partial^2 U^2\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)}{\partial r\partial s}F_k(r,s)dsdr - E^2(\hat{F}_k(x,y))\\
&= \int_{R^2} \frac{\partial^2 U^2(r,s)}{\partial r\partial s}F_k(x-h_nr,y-h_ns)dsdr - E^2(\hat{F}_k(x,y))\\
&= F_k(x,y) - h_nF_{k(1)}(x,y)\int_{R^2} r\frac{\partial^2 U^2(r,s)}{\partial r\partial s}dsdr\\
&\qquad -h_nF_{k(2)}(x,y)\int_{R^2} s\frac{\partial^2 U^2(r,s)}{\partial r\partial s}dsdr + o(h_n) - E^2(\hat{F}_k(x,y))\\
&= F_k(x,y) - F^2(x,y) + O(h_n),
\end{aligned}
\tag{3.4}
$$

where $F_{k(1)}(x,y)$ and $F_{k(2)}(x,y)$ are the first-order partial derivatives with respect to the first and second components, respectively, at point (x, y).

Also, we have for $j \neq 1$

$$
\begin{aligned}
&\operatorname{Cov}\left(U\left(\frac{x-X_1}{h_n},\frac{y-X_{k+1}}{h_n}\right),U\left(\frac{x-X_j}{h_n},\frac{y-X_{k+j}}{h_n}\right)\right)\\
&= E\left(U\left(\frac{x-X_1}{h_n},\frac{y-X_{k+1}}{h_n}\right)U\left(\frac{x-X_j}{h_n},\frac{y-X_{k+j}}{h_n}\right)\right) - E^2(\hat{F}_k(x,y))\\
&= \int_{R^4} U\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)U\left(\frac{x-w}{h_n},\frac{y-v}{h_n}\right)dF_{kj}(r,s,w,v) - E^2(\hat{F}_k(x,y))\\
&= \int_{R^4} \frac{\partial^2 U\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)}{\partial r\partial s}\frac{\partial^2 U\left(\frac{x-w}{h_n},\frac{y-v}{h_n}\right)}{\partial w\partial v}F_{kj}(r,s,w,v)dsdrdvdw - E^2(\hat{F}_k(x,y))
\end{aligned}
\tag{3.5}
$$

$$
\begin{aligned}
&= \int_{R^4} u\left(\frac{x-r}{h_n},\frac{y-s}{h_n}\right)u\left(\frac{x-w}{h_n},\frac{y-v}{h_n}\right)F_{kj}(r,s,w,v)dsdrdvdw - E^2(\hat{F}_k(x,y))\\
&= F_{kj}(x,y,x,y) - F^2(x,y) + O(h_n^2).
\end{aligned}
\tag{3.6}
$$

So the proof is complete. □

Proposition 3.1. *Let A1, A2 and A3 be satisfied. Then,*

$$(n-k)MSE(\hat{F}_k(x,y)) = F_k(x,y) - F_k^2(x,y) + 2\sum_{j=2}^{\infty}(F_{kj}(x,y,x,y) - F_k^2(x,y))$$
$$+ O(nh_n^4 + h_n) + a_n,$$

where a_n is independent of h_n and tends to 0, as $n \to \infty$. Therefore, an optimal convergence rate of the MSE is achieved by choosing $h_n = O(n^{-1/3})$.

Proof. For every $x, y \in R$,

$$\mathrm{MSE}(\hat{F}_k(x,y)) = E(\hat{F}_k(x,y) - F_k(x,y))^2 = \mathrm{Bias}^2\left(\hat{F}_k(x,y)\right) + \mathrm{Var}\left(\hat{F}_k(x,y)\right). \tag{3.7}$$

We can write the last term in (3.7) as

$$\mathrm{Var}\left(\hat{F}_k(x,y)\right) = \frac{1}{n-k}\mathrm{Var}\left(U\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right)\right)$$
$$+ \frac{1}{(n-k)^2}\sum_{j\neq 1}\mathrm{Cov}\left(U\left(\frac{x-X_1}{h_n}, \frac{y-X_{k+1}}{h_n}\right), U\left(\frac{x-X_j}{h_n}, \frac{y-X_{k+j}}{h_n}\right)\right).$$

Using Lemma 3.1, we have

$$\begin{aligned}(n-k)\mathrm{MSE}(\hat{F}_k(x,y)) = F_k(x,y) - F_k^2(x,y) + 2\sum_{j=2}^{\infty}(F_{kj}(x,y,x,y) - F_k^2(x,y)) \\ +O(nh_n^4 + h_n) + a_n,\end{aligned} \tag{3.8}$$

where

$$a_n = \frac{2}{n-k}\sum_{j=2}^{n-k}(j-1)(F_{kj}(x,y,x,y) - F_k^2(x,y)) + 2\sum_{j=n-k+1}^{\infty}(F_{kj}(x,y,x,y) - F_k^2(x,y)).$$

The expression a_n is independent of h_n and tends to 0.

Finally, by optimizing the right-hand side of (3.8) with respect to h_n, the optimal bandwidth rate is of order $n^{-1/3}$, so the proof is complete. □

4. Simulation study

Now, we compare the performance of the kernel and histogram estimators via a simulation study for multivariate normal sequences using R software. The sequence $\{X_n, n \geq 1\}$ is called a normal sequence if for $n \geq 2$, the random vector $(X_1, ..., X_n)$ has the multivariate normal distribution. A multivariate normal distribution is NSD if the off-diagonal elements of its covariance matrix are non positive (Hu, 2000). Thus, for generating the NSD data, we suppose that $X_1, ..., X_n$ have multivariate normal distribution with a zero mean vector and covariance matrix

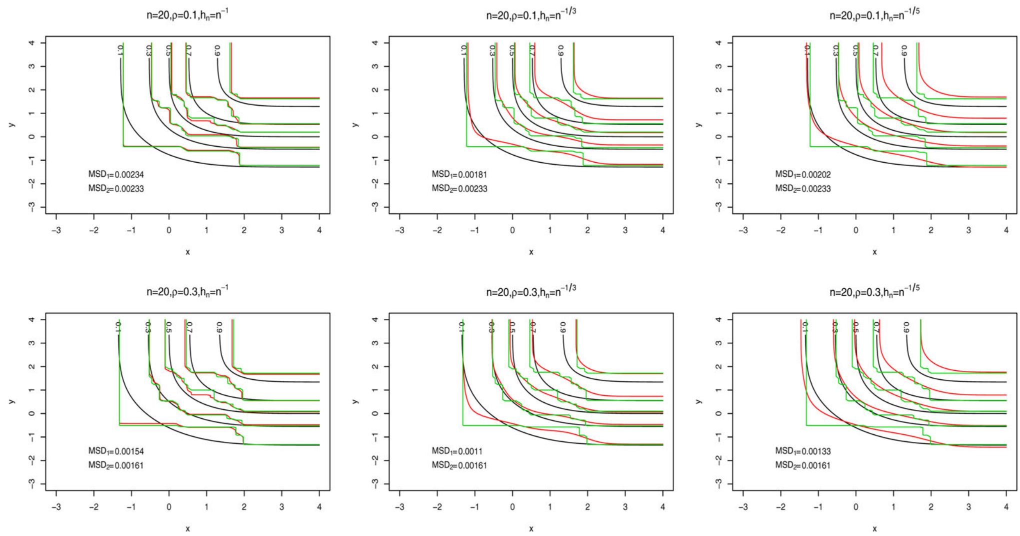

Figure 1. Contour plots of the bivariate normal distribution function (black), histogram estimator (green) and kernel estimator (red) of $F_1(x, y)$ for sample size $n = 20$.

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

where $\rho > 0$. For $n = 20$, 100, we generate one sample from the n-dimensional multivariate normal distribution with $\rho = 0.1, 0.3$. Note that, if $\rho > 0.33$ for some n, Σ is numerically not positive definite, so we choose $\rho = 0.3$ as a strong dependence. Then, for $k = 1$, 2, we compute the kernel and histogram estimators using $h_n = n^{-1}, n^{-1/3}, n^{-1/5}$ and $U(\cdot, \cdot)$ as the bivariate normal distribution with a zero mean vector and covariance matrix

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

Results for $k = 1$, 2 and different values of n, ρ and h_n are presented in Figures 1–4, respectively. Also for simplicity of comparison, we compute the following mean square distances (MSDs) between estimators and $F_k(x, y)$ for all x, y:

$$\mathrm{MSD}_1 = \frac{1}{N}\sum_{x,y}\left(\hat{F}_k(x,y) - F_k(x,y)\right)^2,$$

$$\mathrm{MSD}_2 = \frac{1}{N}\sum_{x,y}\left(\tilde{F}_k(x,y) - F_k(x,y)\right)^2,$$

where N is the product of all numbers r and s.

Figures 1 and 2 show that for $F_1(x, y)$:

- When n is small ($n = 20$), we don't have a good fit but the mean square distance of the kernel estimator is better than the histogram estimator for all h_n and ρ.

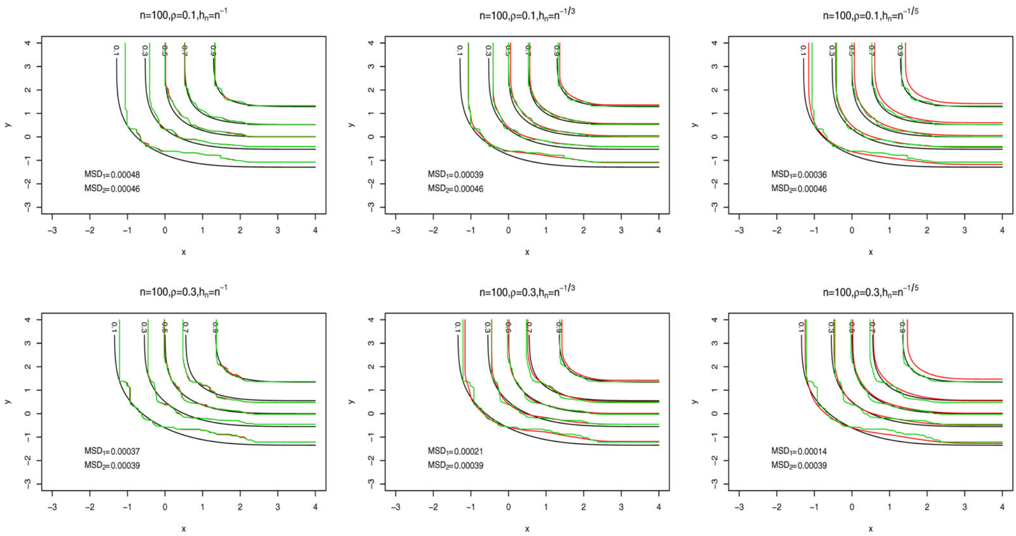

Figure 2. Contour plots of the bivariate normal distribution function (black), histogram estimator (green) and kernel estimator (red) of $F_1(x,y)$ for sample size $n=100$.

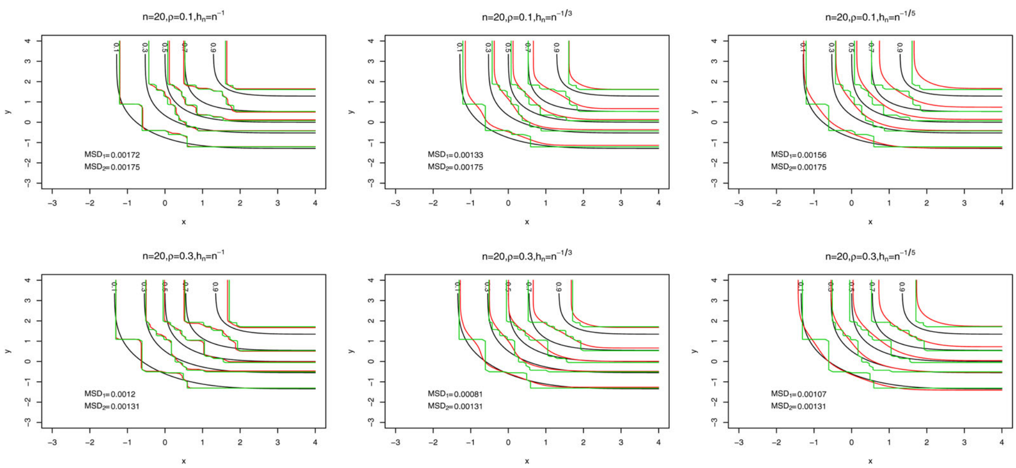

Figure 3. Contour plots of the bivariate normal distribution function (black), histogram estimator (green) and kernel estimator (red) of $F_2(x,y)$ for sample size $n=20$.

- The mean square distance of the kernel estimator is less than for the histogram estimator for all cases.
- When $n=100$, the mean square distance of both estimators is decreased and the histogram estimator is closed to that of kernel estimator for two dependence cases.
- For all n and ρ, the best choice for the bandwidth rate is $h_n=n^{-1/3}$.
- For all n and h_n if ρ is increased, the mean square distance of both estimators is decreased but the kernel estimator is significantly better than the histogram estimator.

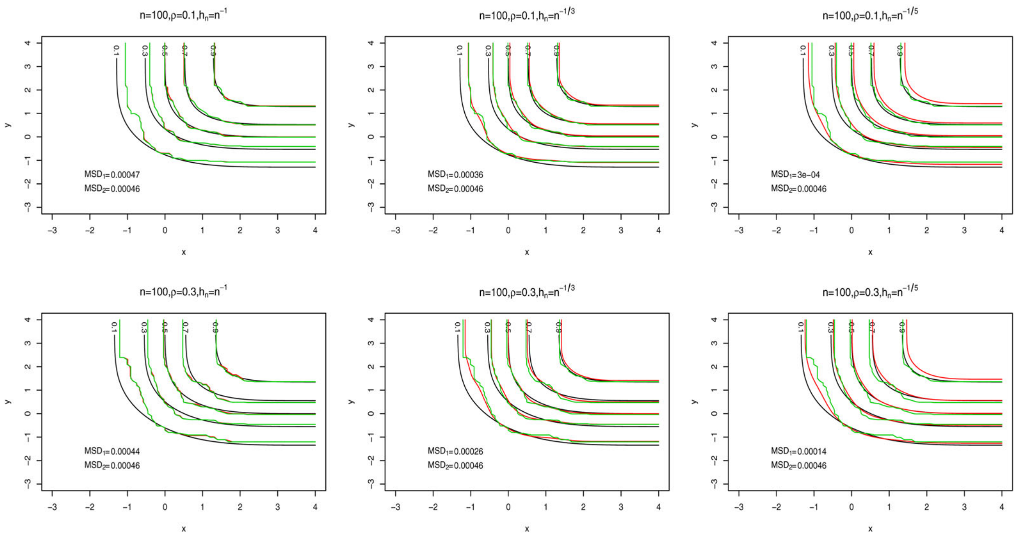

Figure 4. Contour plots of the bivariate normal distribution function (black), histogram estimator (green) and kernel estimator (red) of $F_2(x, y)$ for sample size $n = 100$.

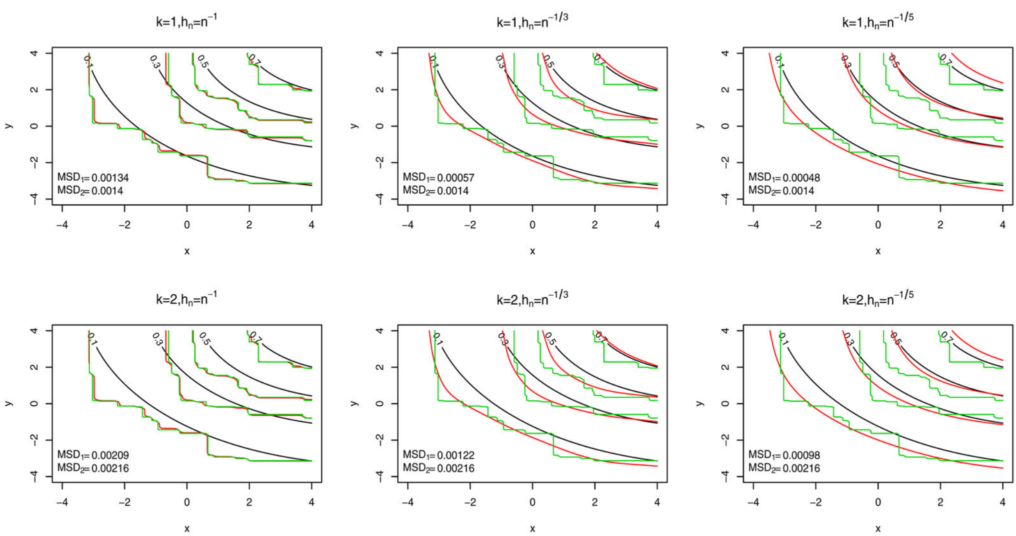

Figure 5. Contour plots of the bivariate normal distribution function (black), histogram estimator (green) and kernel estimator (red) of $F_k(x, y)$ for the differences of annual total rainfall depth series.

Figures 3 and 4 show that for $F_2(x, y)$:

- When n is small, we don't have a good fit but the mean square distance of the kernel estimator is better than the histogram estimator for all cases.
- For all cases, the mean square distance of the kernel estimator is less than the histogram estimator.
- When n is large, the difference between estimators is very small.
- For all n and ρ, the best choice for bandwidth rate is $h_n = n^{-1/3}$.

5 Application on hydrology data

As a real world example, we consider the annual total rainfall depth in the Paraopeba River catchment (Brazil) for years 1950–51 to 1998–99. This historical time series has been considered in Kheyri et al. (2019a) as an applied example for estimation of the distribution function under NSD dependence. They showed that the difference of real data has NSD property.

Here, we use this example and for bandwidth rates $h_n = n^{-1}, n^{-1/3}, n^{-1/5}$, for $k = 1$, 2, compute the kernel estimator $\hat{F}_k$ using $U(\cdot,\cdot)$ as the bivariate normal distribution. We summarize the results in Figure 5. In addition, for simplicity of comparison, we quantify the mean square distances between the estimators and true distribution function of the bivariate normal. In Figure 5, we see that the bandwidth rate $h_n = n^{-1/3}$ is considerably better than $h_n = n^{-1}$ and the results for bandwidth rates $h_n = n^{-1/3}$ and $h_n = n^{-1/5}$ are almost similar. So, for all desired values of h_n, the kernel estimator behaves better than the histogram estimator.

6. Conclusion

In this paper, we have discussed the kernel estimation of the bivariate distribution function under negative superadditive dependence. We derived sufficient conditions in order to prove the exponential inequalities for uniform convergence, which generalized and improved the corresponding ones for NA random variables. We also proved the convergence rate for the kernel estimator of the distribution function that is of the order $O(\sqrt{h_n^4 \log{(n-k)}/(n-k)})$. Furthermore, we obtained the optimal bandwidth convergence rate, which is of the order $n^{-1/3}$. We compared the kernel and histogram estimators in a simulation study, and applied the results to a real world example in hydrology.

Acknowledgements

The authors would like to thank the editor and anonymous referees for their valuable comments and suggestions that improved the quality of the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research has been supported by a grant from Ferdowsi University of Mashhad (No.2/52503).

ORCID

M. Amini http://orcid.org/0000-0002-8336-201X
H. Jabbari http://orcid.org/0000-0002-7028-9052

References

Azevedo, C., and P. E. Oliveira. 2000. Kernel-type estimation of bivariate distribution function for associated random variables. *New Trends in Probability and Statistics* Vol. 5, Proceedings of the 6th Tartu Conference, VSP, 17–25.

Chen, M., H. Chen, Z. Wang, Z. Lu, and X. Wang. 2020. Complete moment convergence for partial sums of arrays of rowwise negatively superadditive dependent random variables. *Communications in Statistics - Theory and Methods* 49 (5):1158–73. doi:10.1080/03610926.2018.1554136.

Christofides, T. C., and E. Vaggelatou. 2004. A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis* 88 (1):138–51. doi:10.1016/S0047-259X(03)00064-2.

Cong, S. T., M. Tran, and D. V. Le. 2020. On the almost sure convergence for sums of negatively superadditive dependent random vectors in Hilbert spaces and its application. *Communications in Statistics - Theory and Methods* 49 (11):2770–86. doi:10.1080/03610926.2019.1584304.

Donsker, M. D. 1951. An invariance principle for certain probability limit theorems. *Memoirs of the American Mathematical Society* 6: 1–12.

Henriques, C., and P. E. Oliveira. 2003. Estimation of a two dimensional distribution function under association. *Journal of Statistical Planning and Inference* 113 (1):137–50. doi:10.1016/S0378-3758(01)00296-8.

Henriques, C., and P. E. Oliveira. 2008. Strong convergence rates for the estimation of a covariance operator for associated samples. *Statistical Inference for Stochastic Processes* 11 (1):77–91. doi:10.1007/s11203-006-9007-3.

Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* 58 (301):13–30. doi:10.1080/01621459.1963.10500830.

Hu, T. Z. 2000. Negatively superadditive dependence of random variables with applications. *Chinese Journal of Applied Probability Statistics* 16:133–44.

Jabbari, H. 2009. Almost sure convergence of kernel bivariate distribution function estimator under negative association. *Journal of Statistical Research of Iran* 6:243–55.

Jabbari, H., and H. A. Azarnoosh. 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.

Kheyri, A., M. Amini, H. Jabbari, and A. Bozorgnia. 2019a. Convergence rate for estimator of distribution function under NSD assumption with an application. *Journal of the Iranian Statistical Society* 18 (2):21–37. doi:10.29252/jirss.18.2.21.

Kheyri, A., M. Amini, H. Jabbari, and A. Bozorgnia. 2019b. Kernel density estimation under negative superadditive dependence and its application for real data. *Journal of Statistical Computation and Simulation* 89 (12):2373–92. doi:10.1080/00949655.2019.1619738.