

Convergence Rate for Estimator of Distribution Function under NSD Assumption with an Application

Azam Kheyri ¹, Mohammad Amini ¹, Hadi Jabbari ¹, and Abolghasem Bozorgnia ²

¹ Department of Statistics, Ordered and Spatial Data Center of Excellence, Ferdowsi University of Mashhad, Mashhad, Iran.

² Department of Statistics, Khayyam University, Mashhad, Iran.

Received: 16/04/2018, Revision received: 26/12/2018, Published online: 29/08/2019

Abstract. In this paper, the kernel distribution function estimator for negative super-additive dependent (NSD) random variables is studied. The exponential inequalities and exponential rate for the kernel estimator are investigated. Under certain regularity conditions, the optimal bandwidth is determined using the mean squared error and is found to be the same as that in the independent identically distributed case. A simulation study to examine the behavior of the kernel and empirical estimators is given. Moreover, a real data set in hydrology is analyzed to demonstrate the structure of negative superadditive dependence, and as a result, the kernel distribution function estimator of the data is investigated.

Keywords. Exponential Rates, Kernel Estimation, Negative Superadditive Dependence.

MSC: 62G07; 62G20.

Azam Kheyri (azam.kheyri@mail.um.ac.ir)
Corresponding Author: Mohammad Amini (m-amini@um.ac.ir)
Hadi Jabbari (Jabbarinh@um.ac.ir)
Abolghasem Bozorgnia (a.bozorgnia@khayyam.ac.ir)

1 Introduction

Estimation of distribution functions has been one of the classical problems in statistics. There has been extensive work on statistical estimation of a distribution function using kernel methods. Several properties of the kernel distribution function estimator have been investigated by many authors over the years. In the case of independent identically distributed random variables, Nadaraya (1964), Winter (1973) and Yamato (1973) proved almost uniform convergence of the kernel estimator; Watson and Leadbetter (1964) established asymptotic normality for the kernel estimator. Azzalini (1981) obtained an asymptotic expression for the MSE of the kernel estimator, and he as well as Lejeune and Sarda (1992), and Sarda (1993) derived the asymptotically optimal (in the MSE sense) smoothing parameter h_n . Swanepoel (1988) proved that the uniform kernel estimator is optimal (in the mean integrated squared error sense) and derived an expression for the smoothing parameter h_n , which minimizes the mean integrated squared error.

All references cited above deal with the *i.i.d.* case. In a Markovian framework, Roussas (1969) proved almost sure uniform convergence of the kernel estimator. Under various modes of mixing, Cai and Roussas (1992), and Cai (1993) established almost sure rates of uniform convergence for the kernel estimator. Roussas (1995) obtained the asymptotic normality of the kernel estimator under association and in the framework of random fields.

However, despite the vast amount of reported studies in the literature on properties of the kernel estimator under dependent cases, there are no reported studies concerning NSD random variables introduced by Hu (2000). The definition of NSD random variables is expressed on the basis of the superadditive functions. A function $\phi : R^n \rightarrow R$ 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, \dots, X_n) is said to be NSD if

$$E\phi(X_1, \dots, X_n) \leq E\phi(X_1^*, \dots, X_n^*), \quad (1.1)$$

where X_1^*, \dots, 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. Also, a sequence $\{X_n, n \geq 1\}$ of random variables is NSD if every finite subfamily is NSD.

In this paper, we consider samples that satisfy the notion of NSD. Christofiedes 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. A number of limit theorems for NSD random variables have been studied; some recent works are Meng *et al.* (2017) and Wang *et al.* (2018a,b).

The remaining sections of the paper are organized as follows. The exponential inequality for the uniform convergence rate of the kernel estimator under NSD are introduced in the next section where we derive the convergence rate of the kernel estimator. The asymptotic properties and convergence rate of the mean square error are studied in Section 3, and then the optimal choice of the bandwidth h_n is determined. Moreover, we illustrate the behavior of the kernel and empirical estimators with respect to their empirical the mean square error (EMSE) in Section 4. As an application, in Section 5, rainfall depth data is considered and after checking the dependency structure, the kernel estimator of the corresponding distribution function is obtained.

2 An Exponential Convergence Rate

Let $\{X_n; n \geq 1\}$ be a sequence of random variables with the common unknown distribution function $F(\cdot)$ and probability density function $f(\cdot)$. The most common estimator of $F(\cdot)$, constructed on the basis of the first n random variables from the sequence, is the empirical distribution function, $F_n(\cdot)$ defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x). \tag{2.1}$$

This estimator does not take into consideration the smoothness of $F(\cdot)$, this omission may be improved by utilizing the following kernel estimator $\hat{F}_n(\cdot)$, as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^n K\left(\frac{x - X_j}{h_n}\right), \tag{2.2}$$

where $K(\cdot)$ is a fixed distribution function and h_n is a sequence of nonnegative real numbers converging to zero.

In this section, we prove an exponential probability inequality for the centered kernel estimator of $F(\cdot)$ and using this inequality, we obtain the exponential convergence

rate for the kernel estimator.

All results are derived under the basic assumption of NSD. The remaining assumptions that we need to prove the main results are listed below.

- A1(i): $\{X_n, n \geq 1\}$ is an identically distributed sequence of NSD random variables with distribution function $F(\cdot)$ and bounded density function $f(\cdot)$.

A1(ii): $f(\cdot)$ has bounded and continuous derivative.

A1(iii): $F_{ij}(\cdot, \cdot)$ (the joint distribution function of (X_i, X_j)), has all partial derivatives of order 2, and they are all bounded.

- A2(i): $K(\cdot)$ is a distribution function over an interval $I \subseteq R$ with density function $k(\cdot)$.

A2(ii): $\int_R uk(u)du = 0$, and $\mu_2(k) = \int_R u^2k(u)du < \infty$.

- A3: The sequence of bandwidths $\{h_n; n > 1\}$ is such that, as $n \rightarrow \infty$:

$$(i) 0 < h_n \rightarrow 0 \quad (ii) nh_n \rightarrow \infty \quad (iii) nh_n^4 \rightarrow 0$$

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

$$\left| K\left(\frac{x - X_i}{h_n}\right) - E\left[K\left(\frac{x - X_i}{h_n}\right)\right] \right| \leq ch_n \quad a.s. \quad (2.3)$$

Remark 1. Assumptions A1 to A3 are often applied in the asymptotic theory of kernel estimators in the literature. A4 is reasonable because for any NSD sequence of random variables we have

$$K\left(\frac{x - X_i}{h_n}\right) - E\left[K\left(\frac{x - X_i}{h_n}\right)\right] = \int_{-\infty}^{\frac{x - X_i}{h_n}} k(u)du - \int_{-\infty}^{\infty} K\left(\frac{x - v}{h_n}\right)dF(v),$$

where $k(\cdot)$ is the probability density function associated with $K(\cdot)$. Letting $u = x + h_n r$ and $v = x - h_n s$ in the above equation and some calculations, we'll have

$$\left| K\left(\frac{x - X_i}{h_n}\right) - E\left[K\left(\frac{x - X_i}{h_n}\right)\right] \right| \leq h_n \int_{-\infty}^{\infty} k(x + h_n r)dr + h_n \int_{-\infty}^{\infty} K(s)ds = O(h_n). \quad a.s.$$

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(i) and A4 hold true, then for every $\varepsilon > 0$,*

$$P\left(\sup_{x \in R} |\hat{F}_n(x) - E[\hat{F}_n(x)]| > \varepsilon\right) \leq 2 \exp\left(-\frac{n\varepsilon^2}{2c^2h_n^2}\right). \tag{2.4}$$

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

$$T_{i,n}(x) = K\left(\frac{x - X_i}{h_n}\right) - E\left[K\left(\frac{x - X_i}{h_n}\right)\right],$$

then for every $\varepsilon > 0$,

$$\begin{aligned} P\left(\sup_{x \in R} |\hat{F}_n(x) - E[\hat{F}_n(x)]| > \varepsilon\right) &= P\left(\sup_{x \in R} \left|\sum_{i=1}^n T_{i,n}(x)\right| > n\varepsilon\right) \\ &= P\left(\sup_{x \in R} \left[\left(\sum_{i=1}^n T_{i,n}(x)\right)^+ + \left(\sum_{i=1}^n T_{i,n}(x)\right)^-\right] > n\varepsilon\right) \\ &\leq P\left(\sup_{x \in R} \left(\sum_{i=1}^n T_{i,n}(x)\right)^+ > \frac{n\varepsilon}{2}\right) + P\left(\sup_{x \in R} \left(\sum_{i=1}^n T_{i,n}(x)\right)^- > \frac{n\varepsilon}{2}\right) \\ &\leq P\left(\sup_{x \in R} \sum_{i=1}^n T_{i,n}^+(x) > \frac{n\varepsilon}{2}\right) + P\left(\sup_{x \in R} \sum_{i=1}^n T_{i,n}^-(x) > \frac{n\varepsilon}{2}\right), \end{aligned} \tag{2.5}$$

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

$$\begin{aligned} P\left(\sup_{x \in R} \sum_{i=1}^n T_{i,n}^+(x) > \frac{n\varepsilon}{2}\right) &\leq P\left(\sum_{i=1}^n \sup_{x \in R} T_{i,n}^+(x) > \frac{n\varepsilon}{2}\right) \\ &\leq e^{-\frac{nt}{2}} E\left(\exp\left(t \sum_{i=1}^n \sup_{x \in R} T_{i,n}^+(x)\right)\right). \end{aligned} \tag{2.6}$$

If $\phi(x_1, \dots, x_n) = \exp(t \sum_{i=1}^n \sup_{x \in \mathbb{R}} T_{i,n}^+(x))$, then we can show that $\frac{\partial^2 \phi(x_1, \dots, x_n)}{\partial x_i \partial x_j} \geq 0$, for all $i \neq j$. Therefore $\phi(x_1, \dots, x_n)$ is a superadditive function; so using (1.1) and by some calculations, it follows that

$$E \left(\exp \left(t \sum_{i=1}^n \sup_{x \in \mathbb{R}} T_{i,n}^+(x) \right) \right) \leq \prod_{i=1}^n E \left(\exp \left(t \sup_{x \in \mathbb{R}} T_{i,n}^+(x) \right) \right). \quad (2.7)$$

Consequently by (2.6) and (2.7),

$$P \left(\sup_{x \in \mathbb{R}} \sum_{i=1}^n T_{i,n}^+(x) > \frac{n\varepsilon}{2} \right) \leq \exp \left(\frac{-n\varepsilon t}{2} \right) \prod_{i=1}^n E \left(\exp \left(t \sup_{x \in \mathbb{R}} T_{i,n}^+(x) \right) \right). \quad (2.8)$$

Also using Lemma 2.1 under the assumption A4,

$$E \left(\exp \left(t \sup_{x \in \mathbb{R}} T_{i,n}^+(x) \right) \right) \leq \exp \left(-t E \left(\sup_{x \in \mathbb{R}} T_{i,n}^+(x) \right) + \frac{t^2 c^2 h_n^2}{8} \right), \quad (2.9)$$

by (2.8), (2.9) and the fact that $E(\sup_{x \in \mathbb{R}} T_{i,n}^+(x))$ is nonnegative we have,

$$P \left(\sup_{x \in \mathbb{R}} \sum_{i=1}^n T_{i,n}^+(x) > \frac{n\varepsilon}{2} \right) \leq \exp \left(\frac{-n\varepsilon t}{2} + \frac{nt^2 c^2 h_n^2}{8} \right). \quad (2.10)$$

Now substituting the optimal bound into (2.10), we obtain that

$$P \left(\sup_{x \in \mathbb{R}} \sum_{i=1}^n T_{i,n}^+(x) > \frac{n\varepsilon}{2} \right) \leq \exp \left(-\frac{n\varepsilon^2}{2c^2 h_n^2} \right). \quad (2.11)$$

Similarly, an optimal bound for the last term in (2.5) is achieved by

$$P \left(\sup_{x \in \mathbb{R}} \sum_{i=1}^n T_{i,n}^-(x) > \frac{n\varepsilon}{2} \right) \leq \exp \left(-\frac{n\varepsilon^2}{2c^2 h_n^2} \right). \quad (2.12)$$

These complete the proof. \square

In fact, we derived some 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

$$\varepsilon_n^2 = \frac{\alpha h_n^2 \log n}{n}. \quad (2.13)$$

In order to obtain a convergent series in the right-hand side of (2.4), $\alpha > 0$ in (2.13) must be conveniently chosen (it depends on constants appearing in the inequality). So, the convergence rate of the kernel-type estimator is of order $O(\sqrt{h_n^2 \log n/n})$.

3 The Mean Squared Error

In this section, we study the asymptotic properties and convergence rate of the mean squared error (MSE) of the estimator. From which, we derive the optimal bandwidth and convergence rate of order $n^{-1/3}$.

For the formulation of the next auxiliary result, let for all $i \neq j$;

$$\begin{aligned} I_{nij}(x) &= \text{Cov}\left(K\left(\frac{x - X_i}{h_n}\right), K\left(\frac{x - X_j}{h_n}\right)\right), \\ I_{ij}(x) &= \text{Cov}(I(X_i \leq x), I(X_j \leq x)) = F_{ij}(x, x) - F^2(x). \end{aligned}$$

Lemma 3.1. *Suppose that A1(i)(ii), A2 and A3(i) are satisfied. Then*

$$\text{Bias}(\hat{F}_n(x)) = \frac{1}{2}h_n^2 f'(x) \int_{\mathbb{R}} u^2 k(u) du + o(h_n^2), \tag{3.1}$$

and

$$\text{Var}\left(K\left(\frac{x - X_1}{h_n}\right)\right) = F(x)(1 - F(x)) - h_n f(x) \int_{\mathbb{R}} ub(u) du + o(h_n), \tag{3.2}$$

where $b(u) = 2K(u)k(u)$ and $k(u) = dK(u)/du$.

Also if A1(iii) is satisfied, then

$$I_{nij}(x) = I_{ij}(x) + o(h_n). \tag{3.3}$$

Proof. Using integration by parts and Taylor expansion, the expectation of the kernel

estimator can be derived as

$$\begin{aligned}
E(\hat{F}_n(x)) &= \int_{\mathcal{R}} K\left(\frac{x-u}{h_n}\right) f(u) du \\
&= \int_{\mathcal{R}} k(u) F(x-h_n u) du \\
&= \int_{\mathcal{R}} k(u) [F(x) - h_n u f(x) + \frac{1}{2} h_n^2 u^2 f'(x) + o(h_n^2)] du \\
&= F(x) + \frac{1}{2} h_n^2 f'(x) \int_{\mathcal{R}} u^2 k(u) du + o(h_n^2).
\end{aligned}$$

So, the bias of the kernel estimator can be expressed as

$$Bias(\hat{F}_n(x)) = \frac{1}{2} h_n^2 f'(x) \int_{\mathcal{R}} u^2 k(u) du + o(h_n^2).$$

Similarly, the variance of kernel estimator can be written as

$$\begin{aligned}
Var\left(K\left(\frac{x-X_1}{h_n}\right)\right) &= E\left(K\left(\frac{x-X_1}{h_n}\right)\right)^2 - E^2(\hat{F}_n(x)) \\
&= \int_{\mathcal{R}} K^2\left(\frac{x-u}{h_n}\right) f(u) du - E^2(\hat{F}_n(x)) \\
&= \int_{\mathcal{R}} b(u) F(x-h_n u) du - E^2(\hat{F}_n(x)) \\
&= \int_{\mathcal{R}} b(u) [F(x) - h_n u f(x) + o(h_n)] du - (F(x) + o(h_n))^2 \\
&= F(x) \bar{F}(x) - h_n f(x) \int_{\mathcal{R}} u b(u) du + o(h_n), \tag{3.4}
\end{aligned}$$

and

$$\begin{aligned}
I_{mij}(x) &= E\left(K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{x-X_j}{h_n}\right)\right) - E^2(\hat{F}_n(x)) \\
&= \int_{\mathcal{R}} \int_{\mathcal{R}} K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) dF_{ij}(u, v) dudv - E^2(\hat{F}_n(x)) \\
&= \int_{\mathcal{R}} \int_{\mathcal{R}} k(u) k(v) F_{ij}(x-h_n u, x-h_n v) dudv - E^2(\hat{F}_n(x)) \\
&= F_{ij}(x, x) - F^2(x) + o(h_n) \\
&= I_{ij}(x) + o(h_n), \tag{3.5}
\end{aligned}$$

so the proof is complete. □

Proposition 3.1. *Suppose that A1, A2 and A3 are satisfied. Then,*

$$\begin{aligned} nMSE(\hat{F}_n(x)) &= F(x)\bar{F}(x) + \sum_{i \neq j} I_{ij} - h_n f(x) \int_{\mathbb{R}} b(u) du \\ &\quad + nh_n^4 \left[\frac{f'(x)}{2} \int_{\mathbb{R}} u^2 k(u) du \right]^2 + o(h_n + nh_n^4) + a_n, \end{aligned} \tag{3.6}$$

where a_n is independent of h_n and tends to 0, as $n \rightarrow \infty$. Therefore, an optimal convergence rate of the MSE is achieved by choosing $h_n = cn^{-1/3}$. Also an explicit expression for the optimal bandwidth h_n , in the sense of minimizing the MSE, \hat{h}_n is given by

$$\hat{h}_n(x) = \left(\frac{f(x) \int_{\mathbb{R}} ub(u) du}{n \left[f'(x) \int_{\mathbb{R}} u^2 k(u) du \right]^2} \right)^{1/3}, \tag{3.7}$$

provided that $f'(x) \neq 0$.

Proof. For every $x \in \mathbb{R}$,

$$MSE(\hat{F}_n(x)) = E(\hat{F}_n(x) - F(x))^2 = Bias^2(\hat{F}_n(x)) + Var(\hat{F}_n(x)). \tag{3.8}$$

We can write the last term in (3.8) as

$$Var(\hat{F}_n(x)) = \frac{1}{n} Var\left(K\left(\frac{x - X_1}{h_n}\right)\right) + \frac{1}{n^2} \sum_{i \neq j} I_{nij}. \tag{3.9}$$

Using Lemma 3.1, we have

$$\begin{aligned} nMSE(\hat{F}_n(x)) &= F(x)(1 - F(x)) + \frac{1}{n} \sum_{i \neq j} I_{ij}(x) - h_n f(x) \int_{\mathbb{R}} b(u) du \\ &\quad + nh_n^4 \left[\frac{f'(x)}{2} \int_{\mathbb{R}} u^2 k(u) du \right]^2 + o(h_n + nh_n^4) + a_n, \end{aligned} \tag{3.10}$$

where

$$a_n = a_n(x) = \sum_{i=n}^{\infty} \sum_{j=n, j \neq i}^{\infty} I_{ij}(x).$$

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

Finally, by optimizing the right-hand side of (3.10) with respect to h_n , the optimal bandwidth is obtained as (3.7) and the proof is complete. □

4 Simulation Study

In this section, a comparison study of the kernel and empirical estimators via a simulation study for multivariate normal sequences using R software is given. The sequence $\{X_n, n \geq 1\}$ is called normal sequence if, for $n \geq 2$, the random vector (X_1, \dots, X_n) has the multivariate normal distribution. The multivariate normal distribution is NSD if the off-diagonal elements of its covariance matrix are nonpositive (Hu, 2000). So for generating the NSD data, suppose that X_1, \dots, X_n have a multivariate normal distribution with zero mean vector and the covariance matrix

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

where $\rho > 0$. For $n = 20, 50, 100, 200$, we generate 1000 independent samples from the n -dimensional multivariate normal distribution with $\rho = 0.1, 0.3$. Note that, if $\rho > 0.33$ for some n , Σ is not numerically positive definite, so we choose $\rho = 0.3$ as strong dependence. Then for bandwidth rate $h_n = n^{-1}, n^{-1/2}, n^{-1/3}, n^{-1/4}, n^{-1/5}$, we compute the kernel and empirical estimators. Also, the ratio of $EMSE(\hat{F}_n(x))$ (empirical mean square error) to $EMSE(F_n(x))$ is plotted, where $K(\cdot)$ is the distribution function of standard the normal. The results are presented in Figure 1.

Figure 1 shows that:

- For all n considered, $EMSE(\hat{F}_n(x))/EMSE(F_n(x)) < 1$, so the kernel estimator is significantly better than the empirical estimator based on their mean squared error for both weak ($\rho = 0.1$) and strong ($\rho = 0.3$) dependence cases.
- If the sample size n is increased, the MSE of the empirical estimator is close to that of kernel estimator for all two dependence cases.
- If ρ is increased, the MSE of the kernel estimator is better than the empirical estimator for all h_n .
- If h_n is increased, the MSE of the kernel estimator is better than the empirical estimator for all two dependence cases.
- For all n and h_n if ρ is increased, the MSE of the kernel estimator is better than the empirical estimator.

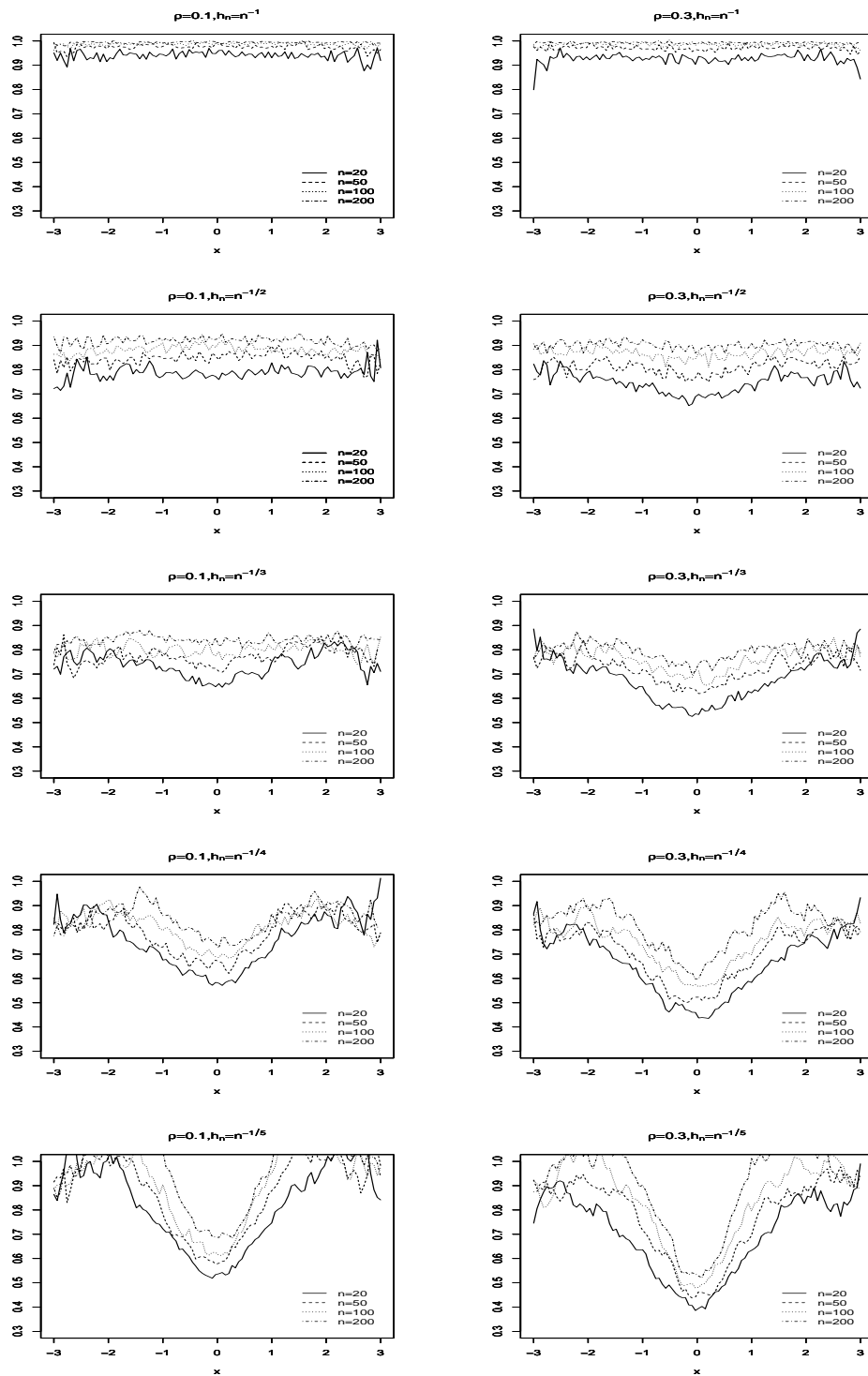


Figure 1: $EMSE(\hat{F}_n(x))/EMSE(F_n(x))$ for $\rho = 0.1, 0.3$ and $n = 20, 50, 100, 200$ in all 1000 generations.

5 Application on Hydrology Data

As a real example, we consider the annual total rainfall depth for the Paraopeba River catchment (Brazil) for water-years 1950-51 to 1998-99. This historical time series has been described in Naghettini (2017). The time series plot is shown in Figure 2(a). Figure 2(b-c) shows the data is not stationary, so for the best modeling, we must transform it.

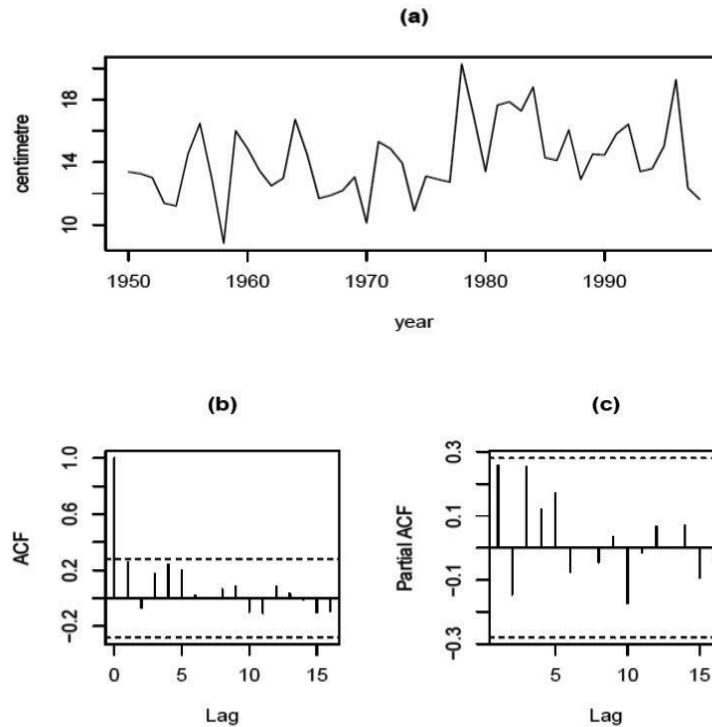


Figure 2: (a) Annual total rainfall depth for the Paraopeba River catchment from 1950-51 through to 1998-99, (b) sample autocorrelation function, (c) sample partial autocorrelation function.

The results of first order difference $y_t = x_t - x_{t-1}$ is displayed in Figure 3. The trend of y_t in Figure 3(a) and its ACF and PACF in Figure 3(b-c) show the stationarity of the time series.

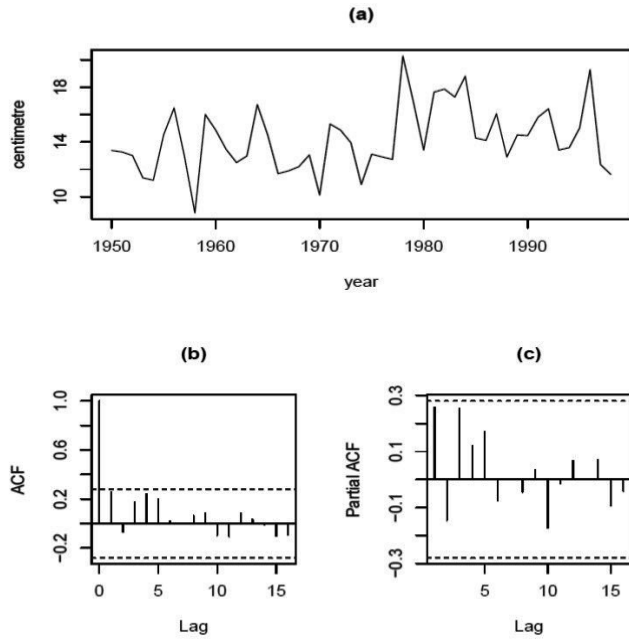


Figure 3: (a) Differences (y_t) of annual total rainfall depth series, (b) sample autocorrelation function, (c) sample partial autocorrelation function.

Then we consider the moving average process $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}$ for $q \geq 1$ using a white noise process Z_t arising from a closer investigation of y_t . Due to the Akaike information criterion (AIC), we expect that the moving average model of order 2 (MA(2)) is suitable for y_t . Finally, MA(2) with $\hat{\theta}_1 = -0.6575$, $\hat{\theta}_2 = -0.3425$, $\hat{\mu}_z = 0$ and $\hat{\sigma}_z^2 = 4.838$ is the best model according to Akaike information criterion (AIC) with $AIC = 223.33$. Testing the normality of the errors is done by Jarque-Bera test (Jarque and Bera, 1980) and the corresponding p-value is equal to 0.2331. The next option is the examination of the existing autocorrelation in residuals. The usual test used in this approach is the Portmanteau test (Ljung and Box, 1978). The results are summarized in Table 1. The null hypothesis of no autocorrelation is not rejected since the p-values are not less than the significance level 0.05.

Table 1: Portmanteau test for lags 5(5)30.

Lags	Statistic	p-value
5	3.415	0.489
10	5.453	0.632
15	7.542	0.720
20	9.436	0.756
25	11.127	0.779
30	13.390	0.753

So, due to the fact that the distributions of linear transformations or linear combination of multivariate normal variables are again multivariate normal (Meng *et al.*, 2017) and according to the above conditions, differences of the real data has a multivariate normal distribution with zero mean vector and covariance matrix $\Sigma = [\sigma_{ij}]$ such that for all $i, j = 1, 2, \dots, n$,

$$\sigma_{ii} = (1 + \theta_1^2 + \theta_2^2)\sigma_z^2,$$

and

$$\sigma_{ij} = \begin{cases} \theta_1(1 + \theta_2) & ; |j - i| = 1 \\ \theta_1 & ; |j - i| = 2 \\ 0 & ; |j - i| > 2 \end{cases}$$

Noting that the multivariate normal distribution is NSD if the off-diagonal elements of its covariance matrix are nonpositive, then for $\hat{\theta}_1 = -0.6575$ and $\hat{\theta}_2 = -0.3425$, we have NSD property in data.

For bandwidth rates $h_n = n^{-1}, n^{-1/3}, n^{-1/5}$, we have computed the kernel estimator \hat{F}_n using $K(\cdot)$ as the standard normal distribution and have summarized the results in Figure 4.

In addition, the goodness of fit test statistics for testing normality regarding some well-know divergence measures based on density is given in Table 2. We denote the test statistics Hellinger, Jeffreys, total variation and chi-squared distance by TH, TJ, TT and T_χ , respectively (see Pardo, 2006). It can be seen from Figure 4 and Table 2 that the bandwidth rate $h_n = n^{-1/5}$ is better than $h_n = n^{-1/3}$ and also bandwidth rate $h_n = n^{-1/3}$ is considerably better than $h_n = n^{-1}$. Note that this does not contradict the optimality of $h_n = n^{-1/3}$, because this optimality is related to the MSE criterion.

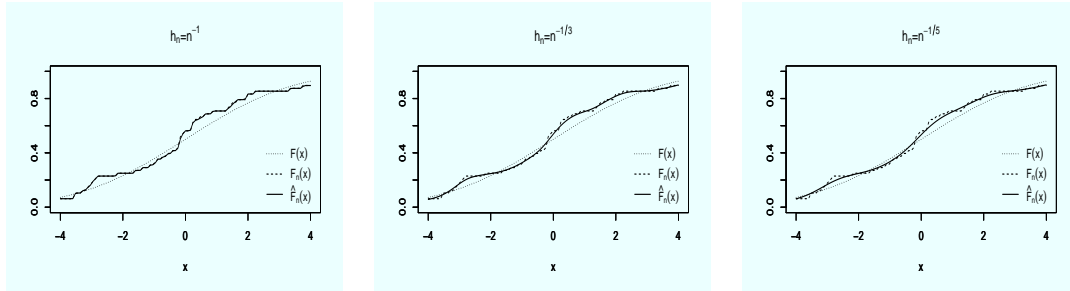


Figure 4: Normal distribution function with zero mean and variance $(1 + \hat{\theta}_1^2 + \hat{\theta}_2^2)\hat{\sigma}_z^2$ and its kernel estimator for the differences of annual total rainfall depth series.

Table 2: The test statistics of normality.

h_n	TH	TJ	TT	T_χ
n^{-1}	0.1653	0.8460	0.7839	0.9378
$n^{-1/3}$	0.0382	0.3187	0.3909	0.6822
$n^{-1/5}$	0.0249	0.2070	0.3249	0.4141

6 Conclusion

In this paper, we have discussed the kernel estimation of distribution function under negative superadditive dependence. We derived some sufficient conditions in order to prove exponential inequalities for uniform convergence, which generalized and improved the corresponding ones for NA random variables. Also, we proved that the convergence rate for the kernel estimator of the distribution function is of order $O(\sqrt{h_n^2 \log n/n})$. Furthermore, the optimal bandwidth convergence rate is obtained, which is of order $n^{-1/3}$. Comparison of the kernel and empirical estimators was discussed in a simulation study. Moreover, a real data set in hydrology was analyzed to demonstrate the structure of negative superadditive dependence for data, and as a result, the kernel distribution function estimator of the data was investigated.

References

- Azzalini, A. (1981), A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326–328.
- Cai, Z. W. (1993), Uniform strong convergence and rates for the kernel estimators of a distribution function and a regression function under weakly dependent assumptions. *Chinese Journal of Applied Probability and Statistics*, **9**, 11–17.
- Cai, Z. W. and Roussas, G. G. (1992), Uniform strong estimation under α -mixing with rates. *Statistics & Probability Letters*, **25**, 211–224.
- Christofiedes, T. C. and Vaggelatou, E. (2004), A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis*, **88**, 138–151.
- Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**(301), 13–30.
- Hu, T. Z. (2000), Negatively superadditive dependence of random variables with applications. *Chinese Journal of Applied Probability Statistics*, **16**, 133–144.
- Jarque, C. M. and Bera, A. K. (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, **6**, 255–259.
- Lejeune, M. and Sarda, P. (1992), Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis*, **14**, 457–471.
- Ljung, G. M. and Box, G. E. P. (1978), On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.
- Roussas, G. G. (1969), Non-parametric estimation of the transition distribution function of a Markov process. *The Annals of Mathematical Statistics*, **40**, 1386–1400.
- Roussas, G. G. (1995), Asymptotic normality of a smooth estimate of the distribution function under association in a random field framework. *Statistics & Probability Letters*, **24**, 77–90.
- Nadaraya, E. A. (1964), Some new estimates for distribution function. *Theory Probability Application*, **9**, 497–500.

- Naghattini, M. (2017), *Fundamentals of statistical hydrology*. Springer International Publishing Switzerland.
- Pardo, L. (2006), *Statistical Inference Based on Divergence Measures*. New York: Chapman and Hall/CRC.
- Sarda, P. (1993), Smoothing parameter selection for smooth distribution functions, *Journal of Statistical Planning and Inference*, **35**, 65–75.
- Swanepoel, J. W. H. (1988), Mean integrated squared error properties and optimal kernels when estimating a distribution function, *Communications in Statistics - Theory and Methods*, **17**, 3785–3799.
- Tong, Y. L. (1990), *The Multivariate normal distribution*. Springer Series in Statistics.
- Meng, B., Wang, D., and Wu, Q. (2017), Complete convergence and complete moment convergence for arrays of rowwise negatively superadditive dependent random variables, *Communications in Statistics–Theory and Methods*, **47**, 3910–3922.
- Wang, X., Wu, Y., and Hu, S. (2018), Strong and weak consistency of LS estimators in the EV regression model with negatively superadditive–dependent errors, *AStA Advances in Statistical Analysis*, **102**, 41–65.
- Wang, X., Wu, Y., and Hu, S. (2018), Complete moment convergence for double-indexed randomly weighted sums and its applications, *Statistics*, **52**, 503–518.
- Watson, G. S. and Leadbetter, M. R. (1964), Hazard analysis II, *Sankhya*, **A(26)**, 101–116.
- Winter, B. B. (1973), Strong uniform consistency of integrals of density estimators, *The Canadian Journal of Statistics*, **1(2)**, 247–253.
- Yamato, H. (1973), Uniform convergence of an estimator of a distribution function, *Bulletin of Mathematics and Statistics*, **15**, 69–78.

