Kernel density estimation under widely orthant dependence
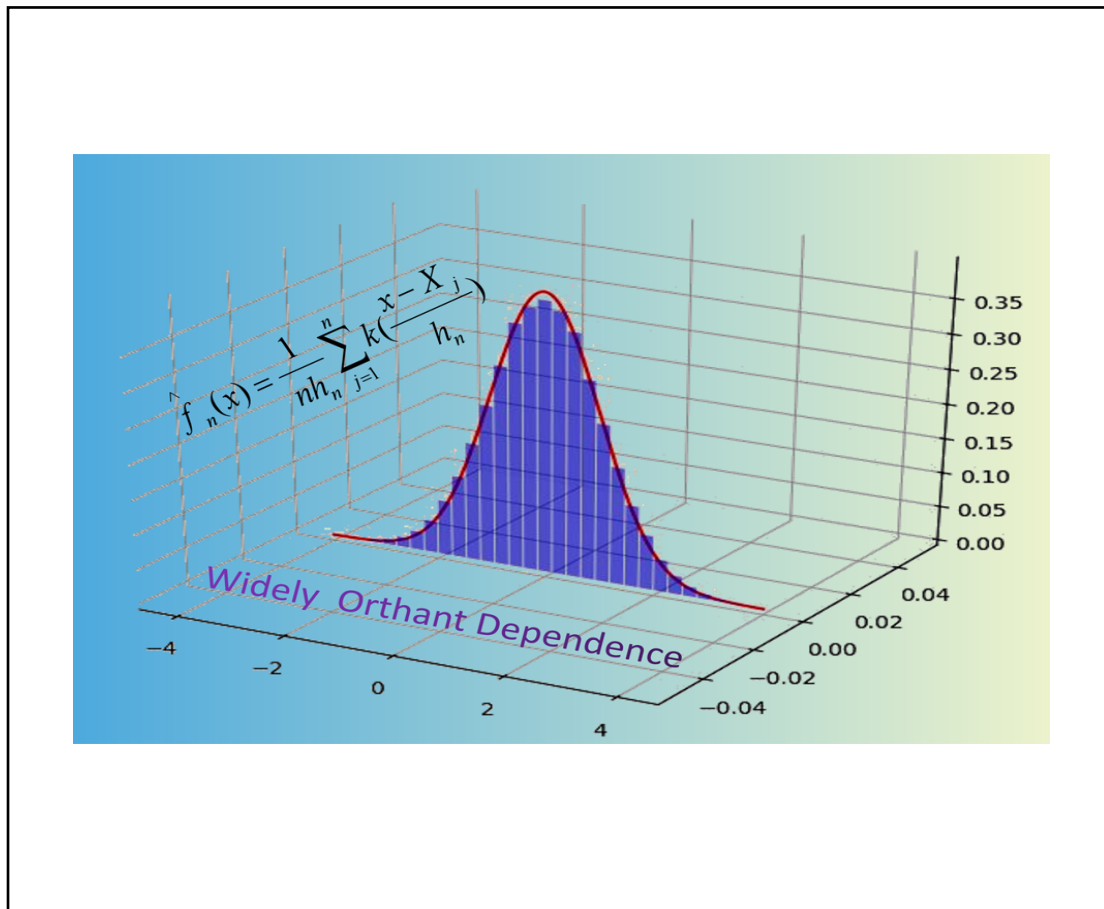
Wei Wang[1], Qingqing Wu[1], and Xufei Tang[2] ✉

[1]School of Big Data and Artificial Intelligence, Chizhou University, Chizhou 247000, China;
[2]School of Mathematics and Statistics, Chaohu University, Hefei 238024, China

✉Correspondence: Xufei Tang, E-mail: xufei_tang@163.com

Graphical abstract

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{j=1}^{n}k\left(\frac{x-X_j}{h_n}\right)$$

We investigate the kernel density estimator for widely orthant dependent random variables and obtain the exponential inequalities and the exponential rate for the estimator of density function with a uniform version over compact sets. The consistency of the estimator is also proved.

Public summary

■ We consider samples that satisfy the notion of WOD, which is the most general one among the dependent random variables.

■ The obtained results extend and complement the results of Kheyri et al.[14] to a much more general case.

■ Some numerical analysis and a real data analysis are presented to support the theoretical results.

Kernel density estimation under widely orthant dependence

Wei Wang[1], Qingqing Wu[1], and Xufei Tang[2] ✉

[1]*School of Big Data and Artificial Intelligence, Chizhou University, Chizhou 247000, China;*
[2]*School of Mathematics and Statistics, Chaohu University, Hefei 238024, China*

✉*Correspondence: Xufei Tang, E-mail: xufei_tang@163.com*

Cite This: *JUSTC*, **2022**, 52(2): 2 (7pp)   Read Online   SI Supporting Information

Abstract: The kernel density estimator for widely orthant dependent random variables is studied. The exponential inequalities and the exponential rate for the estimator of a density function with a uniform version over compact sets are investigated. Further, the consistency of the estimator is proved. The results are generalizations of some existing outcomes for both associated and negatively associated samples. The convergence rate of the kernel density estimator is illustrated via a simulation study. Moreover, a real data analysis is presented.

Keywords: exponential rate; kernel estimation; consistency; widely orthant dependent

CLC number: O211.4   **Document code:** A

2020 Mathematics Subject Classification: 62G05; 62G07

1   Introduction

The density estimation of random variables has always been a classical problem in statistics. With the continuous development of statistics, numerous density estimation theories and methods have been proposed. Among them, kernel density estimation is commonly used to display the shape of a dataset without relying on a parametric model. Parzen[1], Rosenblatt[2], and Silverman[3] have provided early results on kernel density estimation, and since then, much research has been conducted in this area.

In most previous studies, it is assumed that random variables are independent and identically distributed with a common density. However, in many stochastic models and statistical applications, the assumption that random variables are independent is not plausible; therefore, the assumption of dependence is more appropriate. Many useful and interesting results have been obtained for dependent random variables . For instance, Yakowitz[4] showed that the kernel density estimator is asymptotically normal when the sample is obtained from a stationary time series. Bagai et al.[5, 6] proved the pointwise and uniform consistency of the kernel estimator under association. Roussas[7], Masry[8] and Oliveira[9] introduced the consistency results for the kernel estimator under association. Henriques et al.[10] obtained an exponential rate for pointwise and uniform convergence on compact sets with additional conditions for the latter case. Ling et al.[11] proved the strong, uniformly strong, and square consistency of the kernel estimator for negative quadrant dependent variables under some suitable conditions. Su et al.[12] obtained the strong and moment consistency of the kernel estimator under suitable conditions. Wang and Li[13] provided the definitions of generalized consistency for kernel density estimation and obtained several types of generalized consistency of kernel density estimation, and so forth.

Recently, Kheyri et al.[14] studied the kernel density estimat-

or for negatively superadditive dependent (NSD) random variables and obtained the exponential inequalities and exponential rate for NSD random variables. As an application, the corresponding results for Farlie-Gumbel-Morgenstern (FGM) sequences have been studied. It is known that some dependence structures, such as negatively orthant dependence (NOD), extended negative dependence (END) and widely orthant dependence (WOD), are much more general than NSD. Notably, the extensibility of the results of Kheyri et al.[14] to a more general case is an issue that has attracted much interest. Moreover, Kheyri et al.[14] did not consider the consistency of the kernel density estimator. Notwithstanding the vast number of reported studies in the literature on the asymptotic properties of the kernel density estimator under dependent cases, there are no reported studies concerning WOD random variables, which are the most general among the aforementioned dependent random variables. In this study, we consider samples that satisfy the WOD notion and study the exponential inequalities, consistency, and convergence rate of the kernel density estimator. The results obtained in this study extend and complement those of Kheyri et al.[14] to a much more general extent. Moreover, some numerical analysis and a real data analysis are presented to support the theoretical results.

Throughout this study, we always assume that *C* denotes a positive constant that only depends on some given numbers and may vary from one place to another. The bandwidth and density function are denoted by h_n and $f(x)$, respectively, while the limits are taken as $n \to \infty$ unless indicated otherwise. The remainder of this paper is organized as follows. The definitions, necessary notations, and some preliminary lemmas are introduced in Section 2. In Section 3, we prove the exponential inequalities for pointwise convergence and, with an extra condition imposed on the kernel function, uniform convergence on compact sets under WOD. In addition,

we derive the convergence rate of the kernel estimator that is of the order $O(\sqrt{\log(ng(n))/nh_n^2})$, $g(n)$ is the dominating coefficient, which will be discussed later. This convergence rate is slightly slower than the best-known rate for independent cases, and it is of the order $O(\sqrt{\log\log n/(nh_n)})$. In Section 4, we prove some results regarding the consistency of the estimator. Moreover, some numerical analysis and a real data analysis are presented to support the theoretical results in Sections 5 and 6, respectively.

2   Preliminaries

Let $\{X_n, n \geqslant 1\}$ be a sequence of random variables with the common unknown density function $f(\cdot)$.

Definition 2.1. Let $k(\cdot)$ be a fixed probability density function and h_n be a sequence of nonnegative real numbers converging to zero. Then, the kernel estimator of the density function $f(\cdot)$ is, as usual, defined as

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} k\left(\frac{x-X_j}{h_n}\right) \tag{1}$$

which is well known to be asymptotically unbiased, provided that the density function is bounded and continuous. Moreover, the convergence of $E[\hat{f}_n(x)]$ to $f(x)$ is uniform on compact sets under these assumptions on $f(\cdot)$.

The concept of WOD random variables was introduced by Wang et al.[15] as follows:

Definition 2.2. A finite collection of random variables $X_1, X_2, ..., X_n$ is said to be widely upper orthant dependent (WUOD) if there exists a finite real number $g_U(n)$ such that for all finite real numbers x_i, $1 \leqslant i \leqslant n$,

$$P(X_1 > x_1, X_2 > x_2, ..., X_n > x_n) \leqslant g_U(n) \prod_{i=1}^{n} P(X_i > x_i).$$

A finite collection of random variables $X_1, X_2, ..., X_n$ is said to be widely lower orthant dependent (WLOD) if there exists a finite real number $g_L(n)$ such that for all finite real numbers x_i,

$$P(X_1 \leqslant x_1, X_2 \leqslant x_2, ..., X_n \leqslant x_n) \leqslant g_L(n) \prod_{i=1}^{n} P(X_i \leqslant x_i).$$

If $X_1, X_2, ..., X_n$ are both WUOD and WLOD, then $X_1, X_2, ..., X_n$ are WOD with dominating coefficients $g(n) = \max\{g_U(n), g_L(n)\}$. A sequence $\{X_n, n \geqslant 1\}$ of random variables is said to be WOD if every finite subcollection is WOD.

Recall that when $g_L(n) = g_U(n) = M$ for some positive constant M, the random variables $\{X_n, n \geqslant 1\}$ are said to be END. In particular, when $g_L(n) = g_U(n) = 1$ for any $n \geq 1$, the random variables $\{X_n, n \geqslant 1\}$ are said to be NOD or negatively dependent. Joag-Dev and Proschan[16] pointed out that negatively associated (NA) random variables are NOD. Hu[17] introduced the NSD concept and pointed out that NSD implies NOD. From the statements above, we can see that the class of WOD random variables includes END, NOD, NSD, and NA random variables, as well as independent random variables as special cases. In particular, every n-dimensional FGM distri-

bution describes a specific WOD structure. Furthermore, Wang et al.[15] also exemplified that WOD encompasses some positively dependent structures. Hence, it is of great interest to study the limiting behavior of WOD random variables and its applications.

A technical problem arises when dealing with $\hat{f}_n(x)$ for WOD variables. In fact, WOD is only preserved under monotone transformations; that is, in general, the variables $k((x-X_1)/h_n), k((x-X_2)/h_n), ...$ are not WOD. One way to resolve this problem is to assume that the kernel $k(\cdot)$ has bounded variation. Now, we introduce a set of assumptions to prove our main results.

A1(i) $\{X_n, n \geqslant 1\}$ is a strictly stationary sequence of WOD random variables with a bounded and continuous density function $f(\cdot)$.

(ii) $f(\cdot)$ has a bounded and continuous derivative of first and second orders.

A2(i) The kernel function $k(\cdot)$ is a probability density of bounded variation and $\int_R k^2(u)\mathrm{d}u < \infty$; furthermore, if $k(\cdot) = k_1(\cdot) - k_2(\cdot)$ where $k_1(\cdot)$ and $k_2(\cdot)$ are non-decreasing functions such that $k_1'(\cdot)$ and $k_2'(\cdot)$ exist and are integrable.

(ii) $\int_R uk(u)\mathrm{d}u = 0$ and $\mu_2(k) = \int_R u^2 k(u)\mathrm{d}u < \infty$.

A3 The sequence of bandwidth $\{h_n; n > 1\}$ is such that, as $n \to \infty$, $0 < h_n \to 0$ and $nh_n \to \infty$.

Remark 2.1. We state that A1−A3 are highly general assumptions that are frequently used in investigating kernel density estimation, except A2(i). Kheyri et al.[14] first provided some common examples of the kernel function $k(\cdot)$ satisfying A2(i).

Lemma 2.1.[18] Let $\{X_n, n \geqslant 1\}$ be a sequence of WOD random variables.

(i) If $\{f_n(\cdot), n \geqslant 1\}$ are all non-decreasing (or non-increasing), then $\{f_n(X_n), n \geqslant 1\}$ are still WOD.

(ii) For each $n \geqslant 1$ and any real number t,

$$E\exp\left\{t\sum_{i=1}^{n} X_i\right\} \leqslant g(n) \prod_{i=1}^{n} E\exp\{tX_i\} \tag{2}$$

Lemma 2.2.[19] Let X be a centered random variable. If there exist $a, b \in R$ such that $P(a \leqslant X \leqslant b) = 1$, then for every $\lambda > 0$,

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

To obtain the main results, we require the following notations. Given A2(i), define

$$\hat{f}_{1,n}(x) = \frac{1}{nh_n}\sum_{j=1}^{n} k_1\left(\frac{x-X_j}{h_n}\right) \ , \ \ \hat{f}_{2,n}(x) = \frac{1}{nh_n}\sum_{j=1}^{n} k_2\left(\frac{x-X_j}{h_n}\right) \tag{3}$$

such that $\hat{f}_n(x) = \hat{f}_{1,n}(x) - \hat{f}_{2,n}(x)$. For each $n \geqslant 1$, $j = 1, ..., n$ and $q = 1, 2$, let

$$T_{q,j,n} = h_n^{-1}\left[k_q\left(\frac{x-X_j}{h_n}\right) - Ek_q\left(\frac{x-X_j}{h_n}\right)\right] \tag{4}$$

Note that, by Lemma 2.1, we can see that if $(X_1, X_2, ..., X_n)$ is WOD, then $(T_{q,1,n}, ..., T_{q,n,n})$ is also WOD. Finally, for each $q = 1, 2$ and $n \geqslant 1$, we define

$$Z_{q,n} = \sum_{j=1}^{n} T_{q,j,n} \qquad (5)$$

According to (1)–(3), we have $\hat{f}_{q,n}(x) - E[\hat{f}_{q,n}(x)] = (1/n)Z_{q,n}$.

Remark 2.2. If the kernel $k(\cdot)$ satisfies A2(i), the functions $k_1(\cdot)$ and $k_2(\cdot)$ may be chosen bounded such that every random variable $T_{q,j,n}$ is bounded by $2h_n^{-1}\| k_q \|_\infty$, where $\| \cdot \|$ represents the supremum norm. Here, we use Lemma 2.2 to control the moment-generating function of the variables. A simple application of this lemma yields the following upper bounds. The proof is straightforward; therefore, it is omitted.

Lemma 2.3.[14] Let $X_1, ..., X_n$ be random variables, and suppose that A2(i) is satisfied. If $T_{q,j,n}$, $q = 1, 2, j = 1, ..., n$, are defined by Eq.(4), then, for every $\lambda > 0$,

$$\prod_{j=1}^{n} E(e^{(\lambda/n)T_{q,j,n}}) \leq \exp\left(\frac{2\lambda^2\|k_q\|_\infty^2}{nh_n^2}\right), \quad q = 1, 2.$$

3 Exponential inequalities and convergence rate

Now, we prove an exponential probability inequality for $Z_{q,n}$ and thereafter for the centered estimator of $f(\cdot)$. Next, we obtain the exponential convergence rate of the centered estimator of $f(\cdot)$.

Lemma 3.1. Suppose that A1(i) and A2(i) are satisfied. Subsequently, for every $\varepsilon > 0$ and $q = 1, 2$,

$$P\left(\frac{1}{n}|Z_{q,n}| > \varepsilon\right) \leq 2g(n)\exp\left(-\frac{n\varepsilon^2 h_n^2}{8\| k_q \|_\infty^2}\right) \qquad (6)$$

Proof. By Markov's inequality, (2), (5), and Lemma 2.3, we find that for every $\lambda > 0$,

$$P\left(\frac{1}{n}Z_{q,n} > \varepsilon\right) \leq e^{-\lambda\varepsilon}E(e^{(\lambda/n)Z_{q,n}}) =$$

$$e^{-\lambda\varepsilon}E(e^{(\lambda/n)\sum_{j=1}^{n}T_{q,j,n}}) \leq$$

$$g(n)e^{-\lambda\varepsilon}\prod_{j=1}^{n}E(e^{(\lambda/n)T_{q,j,n}}) \leq$$

$$g(n)\exp\left(\frac{2\lambda^2\| k_q^2 \|_\infty^2}{nh_n^2} - \lambda\varepsilon\right) \qquad (7)$$

where for each $n \geq 1, j = 1, ..., n$ and $q = 1, 2$. By optimizing the right-hand side of (7), we have

$$P\left(\frac{1}{n}Z_{q,n} > \varepsilon\right) \leq g(n)\exp\left(-\frac{n\varepsilon^2 h_n^2}{8\| k_q^2 \|_\infty^2}\right).$$

Evidently, (7) is also true for $-(1/n)Z_{q,n}$. Thus, we obtain (6), and the proof is complete.

Here, we extend and improve the results of Henriques and Oliveira[9] and Kheyri et al.[14] for associated random variables and NSD random variables, respectively, to the WOD case. One of our main results is presented in the following proposition:

Proposition 3.1. Suppose that A1(i) and A2(i) are satisfied. Then, for every $\varepsilon > 0$ and sufficiently large n,

$$P(|\hat{f}_n(x) - E[\hat{f}_n(x)]| > \varepsilon) \leq 4g(n)\exp\left(-\frac{n\varepsilon^2 h_n^2}{32C}\right) \qquad (8)$$

where $C = \max\{\|k_1\|_\infty, \|k_2\|_\infty\}$.

Proof. Using Eq. (3), we have

$$\hat{f}_n(x) - E[\hat{f}_n(x)] = (\hat{f}_{1,n}(x) - E[\hat{f}_{1,n}(x)]) - (\hat{f}_{2,n}(x) - E[\hat{f}_{2,n}(x)]) = \frac{1}{n}Z_{1,n} - \frac{1}{n}Z_{2,n}.$$

So for all $\varepsilon > 0$ and applying Lemma 3.1,

$$P(|\hat{f}_n(x) - E[\hat{f}_n(x)]| > \varepsilon) \leq P\left(\frac{1}{n}|Z_{1,n}| > \frac{\varepsilon}{2}\right) + P\left(\frac{1}{n}|Z_{2,n}| > \frac{\varepsilon}{2}\right) \leq$$

$$4g(n)\exp\left(-\frac{n\varepsilon^2 h_n^2}{32C}\right) \qquad (9)$$

This completes the proof.

Now, we use the decomposition of a compact interval to derive a uniform exponential rate for the centered estimator.

Proposition 3.2. Suppose that A1(i) and A2(i) are satisfied, the kernel $k(\cdot)$ is Lipschitzian and $t_n/h_n^2 \to 0$ as $n \to \infty$. Then, for every $\varepsilon > 0$, a sufficiently large n, and each interval $[a,b]$,

$$P\left(\sup_{x\in[a,b]}|\hat{f}_n(x) - E[\hat{f}_n(x)]| > \varepsilon\right) \leq \frac{2g(n)(b-a)}{t_n}\exp\left(-\frac{n\varepsilon^2 h_n^2}{128C}\right) \qquad (10)$$

where $C = \max\{\|k_1\|_\infty, \|k_2\|_\infty\}$.

Proof. Let $[a,b]$ be a fixed interval and decompose $[a,b] = \bigcup_{j=1}^{s_n}[z_{n,j} - t_n, z_{n,j} + t_n]$ into s_n subintervals of length $2t_n$. Then, it is easy to show that

$$\sup_{x\in[a,b]}\left|\hat{f}_n(x) - E[\hat{f}_n(x)]\right| \leq \max_{1\leq j\leq s_n}\left|\hat{f}_n(z_{n,j}) - E[\hat{f}_n(z_{n,j})]\right| + \max_{1\leq j\leq s_n}\sup_{x\in[z_{n,j}-t_n, z_{n,j}+t_n]}\left|\hat{f}_n(x) - \hat{f}_n(z_{n,j}) - E[\hat{f}_n(x) - \hat{f}_n(z_{n,j})]\right|.$$

Because $k(\cdot)$ is Lipschitzian, there exists a constant $\xi > 0$ such that, for all $x \in [z_{n,j} - t_n, z_{n,j} + t_n]$,

$$\left|\hat{f}_n(x) - \hat{f}_n(z_{n,j}) - E[\hat{f}_n(x) - \hat{f}_n(z_{n,j})]\right| \leq 2\frac{\xi|x - z_{n,j}|}{h_n^2} \leq \frac{2\xi t_n}{h_n^2}.$$

Therefore, the condition $t_n/h_n^2 \to 0$ as $n \to \infty$ implies that for all $\varepsilon > 0$,

$$P\left(\sup_{x\in[a,b]}|\hat{f}_n(x) - E[\hat{f}_n(x)]| > \varepsilon\right) \leq$$

$$P\left(\max_{1\leq j\leq s_n}\left|\hat{f}_n(z_{n,j}) - E[\hat{f}_n(z_{n,j})]\right| > \varepsilon - \frac{2\xi t_n}{h_n^2}\right) \leq$$

$$\sum_{j=1}^{s_n}P\left(\left|\hat{f}_n(z_{n,j}) - E[\hat{f}_n(z_{n,j})]\right| > \frac{\varepsilon}{2}\right).$$

Thus, by (8),

$$P\left(\sup_{x\in[a,b]}|\hat{f}_n(x) - E[\hat{f}_n(x)]| > \varepsilon\right) \leq 4g(n)s_n\exp\left(-\frac{n\varepsilon^2 h_n^2}{128C}\right) \qquad (11)$$

and the proof is complete.

We derived some sufficient conditions to prove an exponential rate for the kernel estimator of the density function. To prove the convergence rate, we choose ε depending on n as

$$\varepsilon_n^2 = \frac{\alpha\log(ng(n))}{nh_n^2}.$$

To obtain a convergence series in the right-hand side of (8) and (10), $\alpha > 0$ must be conveniently chosen (it depends on a constant appearing in the inequality). Note that in Proposition 3.2, the assumption $t_n/h_n^2 \to 0$ must be replaced by $t_n/h_n^2 \varepsilon_n \to 0$, and this is also verified by the choice made for ε_n and letting $t_n = h_n^2/n$. Therefore, the convergence rate of the kernel estimator is of the order $O(\sqrt{(\log(ng(n))/(nh_n^2)})$. This convergence rate is slightly slower than the best-known rate for independent cases, which is of the order $O(\sqrt{(\log\log n)/(nh_n)})$. Noting that NSD implies WOD, videlicet $g(n) = 1$, our result extends that of Kheyri et al.[14]. Furthermore, the dominating coefficient $g(n)$ can increase with $g(n) = O(n^\tau)$ for any $\tau \geqslant 0$, whereas the convergence rate remains the same. Even if $g(n)$ increases geometrically, our results remain robust to some extent.

4   Consistency

In this section, we study the consistency of kernel density estimation for the WOD sample. First, we present a lemma that is useful for proving the main results.

Lemma 4.1.[14] Let A1, A2(ii), and A3 be satisfied. Then

$$\text{Bias}(\hat{f}_n(x)) = \frac{1}{2}h_n^2 f''(x)\int_R u^2 k(u)\mathrm{d}u + o(h_n^2) \qquad (12)$$

With Lemma 4.1, we can obtain the following results for weak consistency and the rate of strong consistency of the kernel density estimator.

Theorem 4.1. (Weak consistency) Let Assumptions A1–A3 hold, that is, $g(n) = o(n^\tau)$ for some $\tau \geqslant 0$, $\dfrac{nh_n^2}{\log n} \to \infty$. Then, at each point of continuity x of $f(\cdot)$, the estimator $\hat{f}_n(x)$ is weakly consistent, that is, for each $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\hat{f}_n(x) - f(x)| > \varepsilon) = 0.$$

Proof. It is easy to see that

$$\lim_{n\to\infty} P(|\hat{f}_n(x) - f(x)| > \varepsilon) =$$
$$\lim_{n\to\infty} P(|\hat{f}_n(x) - E\hat{f}_n(x) + E\hat{f}_n(x) - f(x)| > \varepsilon) \leqslant$$
$$\lim_{n\to\infty} P(|\hat{f}_n(x) - E\hat{f}_n(x)| > \varepsilon/2) + \lim_{n\to\infty} P(|E\hat{f}_n(x) - f(x)| > \varepsilon/2).$$

It is sufficient to prove that both summands in the last expression tend to 0 as $n \to \infty$. The first summand is stochastic and the second one is the deterministic component (the bias), both of which tend to 0 as $n \to \infty$. By (8), we have that

$$P(|\hat{f}_n(x) - E\hat{f}_n(x)| > \varepsilon) \leqslant Cn^\tau \exp\left(-\frac{n\varepsilon^2 h_n^2}{32C}\right) \leqslant$$
$$Cn^\tau \exp\left(-(\tau+1)\log(n)\right) \leqslant$$
$$Cn^{-1} \to 0 \qquad (13)$$

Combining (12) and (13), we can easily obtain $\lim_{n\to\infty} P(|\hat{f}_n(x) - f(x)| > \varepsilon) = 0$. Hence, $\hat{f}_n(x)$ is consistent in probability. The proof is complete.

Theorem 4.2. (Strong consistency) Suppose that Assumptions A1–A3 are satisfied; if $\dfrac{nh_n^2}{\log(ng(n))} \to \infty$ as $n \to \infty$, then

$$|\hat{f}_n(x) - f(x)| = O(\sqrt{\log(ng(n))/(nh_n^2)}) + O(h_n^2) \text{ a.s..}$$

Proof. According to (12), we have $|E[\hat{f}_n(x)] - f(x)| = O(h_n^2)$. Hence, it suffices to show that

$$|\hat{f}_n(x) - E[\hat{f}_n(x)]| = O(\sqrt{\log(ng(n))/(nh_n^2)}) \text{ a.s..}$$

By virtue of the Borel-Cantelli lemma, one only needs to prove

$$\sum_{n=1}^{\infty} P\left(|\hat{f}_n(x) - E[\hat{f}_n(x)]| > D\sqrt{\log(ng(n))/(nh_n^2)}\right) < \infty$$

for some $D > 0$. In fact, from (8), we can easily obtain

$$\sum_{n=1}^{\infty} P\left(|\hat{f}_n(x) - E[\hat{f}_n(x)]| > D\sqrt{\log(ng(n))/(nh_n^2)}\right) \leqslant$$
$$C\sum_{n=1}^{\infty} g(n)\exp\left(-\frac{D^2\log(ng(n))}{32C}\right) \leqslant$$
$$C\sum_{n=1}^{\infty} n^{-\frac{D^2}{32C}} g(n)^{1-\frac{D^2}{32C}} < \infty,$$

provided that $D > \sqrt{32C}$. This completes the proof.

5   Numerical simulation study

In this section, we conduct a numerical simulation study using R software to examine the performance of the kernel density estimation with the WOD samples. The simulation is conducted for the following two cases.

Case 1 To generate the WOD random variables, for any fixed $n \geqslant 3$, $(X_1, X_2, \cdots, X_n) \sim N(\mathbf{0}, \Sigma)$, where $\mathbf{0}$ represents zero vector and

$$\Sigma = \begin{pmatrix} 1+\theta^2 & -\theta & 0 & \cdots & 0 & 0 & 0 \\ -\theta & 1+\theta^2 & -\theta & \cdots & 0 & 0 & 0 \\ 0 & -\theta & 1+\theta^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta^2 & -\theta & 0 \\ 0 & 0 & 0 & \cdots & -\theta & 1+\theta^2 & -\theta \\ 0 & 0 & 0 & \cdots & 0 & -\theta & 1+\theta^2 \end{pmatrix}_{n\times n}.$$

Here $0 < \theta < 1$. In the following, we consider $\theta = 0.5$ for simplicity. From Joag-Dev and Proschan[16], it can be seen that $(X_1, X_2, \cdots, X_n)$ is an NA vector, and it is, therefore, a special case of the WOD vector. To examine the performance of the kernel density estimation with the WOD samples, we consider the kernel estimator with $k(u) = (1/\sqrt{2\pi})e^{-u^2/2}$. We consider the fixed bandwidth $h_n = n^{-1/5}$ and the bandwidth based on cross-validation (CV), and we take the sample sizes n as $n = 200, 400, 600$, respectively. We use R software to compute the estimators 500 times to obtain the final values and then compare them with $f(x)$ in Figs. 1 and 2 for different sample sizes.

In Figs. 1 and 2, the black lines represent the true probability density function $f(x)$ and the red points are the estimation values. We can see that the estimator $\hat{f}_n(x)$ tends to $f(x)$ as the sample size n increases. These results are in good agreement with the theoretical results that we establish in this study.

Case 2 Let $\{X_i, i = 1,2\}$ and $\{X_i, i \geqslant 3\}$ be independent of each other, where X_1 and X_2 are possibly valued at $x_{1_1} \leqslant x_{1_2} \leqslant \cdots \leqslant x_{1_N}$ and $\{X_i, i \geqslant 3\}$ is a sequence of mutually independent
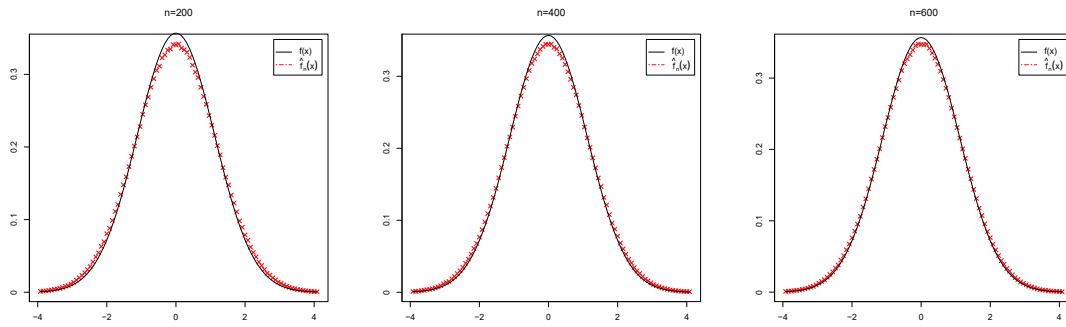
Fig. 1. Comparison of $\hat{f}_n(x)$ and $f(x)$ for the fixed bandwidth $h_n = n^{-1/5}$.
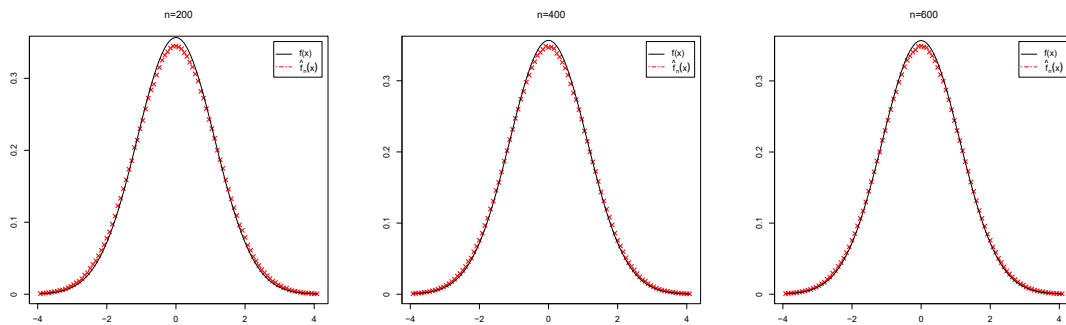

Fig. 2. Comparison of $\hat{f}_n(x)$ and $f(x)$ for the bandwidth based on CV.

random variables. From Liu[20], it follows that $(X_1, X_2, \cdots, X_n)$ is the END vector; therefore, it is WOD. To examine the performance of kernel density estimation with the WOD samples, we generate a sample of the simulations chosen as $X_i \sim \chi^2(3)$, with the sample sizes $n = 100, 200, 400, 800$. Other settings are the same as those in Case 1. For every point $x_i = \dfrac{i}{12.5} - 4$, $i = 1, 2, \cdots, m$, $m = 100$, we also use R software to compute the estimator $N = 500$ times to obtain the global root mean square error (GRMSE), as presented in Table 1 for different sample sizes, where the GRMSE is defined as GRMSE $= \sqrt{\dfrac{1}{mN} \sum_{l=1}^{N} \sum_{i=1}^{m} (\hat{f}_n(x_i, l) - f(x_i, l))^2}$. To express this tendency more clearly, we also obtain the boxplots of $\hat{f}_n(x) - f(x)$ in Fig. 3.

From Table 1, we can see that the GRMSE decreases for

Table 1. The GRMSE of the kernel density estimator.

	$n = 100$	$n = 200$	$n = 400$	$n = 800$
Fixed	0.021705	0.017502	0.015432	0.014240
CV	0.023198	0.018441	0.015190	0.012305

each bandwidth as the sample size increases. For $n \leqslant 200$, the GRMSE with a fixed bandwidth is slightly smaller than that with CV. For $n \geqslant 400$, the GRMSE with CV reduces. Fig. 3 shows the similar results. Overall, the fixed bandwidth and the bandwidth based on CV do not have evident differences.

6　Real data analysis

In this section, we perform a real data analysis for kernel density estimation based on the urbanization rate time series
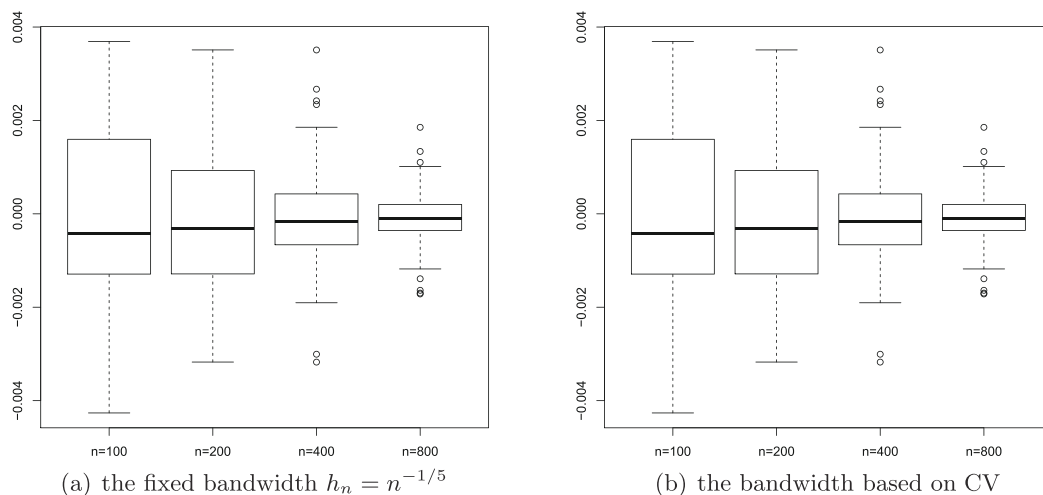

(a) the fixed bandwidth $h_n = n^{-1/5}$　　　(b) the bandwidth based on CV

Fig. 3. Boxplot of $\hat{f}_n(x) - f(x)$ for $n = 100, 200, 400, 800$.

of China (denoted as X_t). Fig. 4 shows 67 yearly urbanization rates from 1949 to 2015. The data are obtained from the website of the National Bureau of Statistics of China. First, we plot the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) in Fig. 4. From Fig. 4, it is evident that the sample is not stationary. For the best modelling, we must transform the data. The results of the first-order difference $y_t = x_t - x_{t-1}$ are shown in Fig. 5. The trend of y_t and its ACF and PACF in Fig. 5 show a stationary process. Thereafter we perform the unit root test and find that this time series is stationary. According to the Akaike information criterion (AIC) and Fig. 5, MA(2) process $Y_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$ with $\hat{\theta}_1 = 0.3075$, $\hat{\theta}_2 = -0.6925$ is the best model, where $\{Z_t, t \geqslant 1\}$ is a Gaussian white noise sequence with $\hat{\mu}_Z = 0$ and $\hat{\sigma}_Z^2 = 0.38$. Apparently, Y_t is normally distributed, with $\hat{\mu}_Y = 0$ and $\hat{\sigma}_Y^2 = (1 + \hat{\theta}_1^2 + \hat{\theta}_2^2)\hat{\sigma}_Z^2 \approx 0.5982$. It is evident that Y_t is not negatively dependent; therefore, it is not NSD. Based on Wang et al.[15], we know that WOD comprises not only negatively dependent structures but also some positively dependent structures. Therefore, it is more reasonable to use WOD to describe the data. Because the distribution of linear transformations or linear combinations of multivariate normal variables are again multivariate normal, the difference in real data has a multivariate normal distribution with a zero mean vector and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \rho_0\sigma_Z^2 & \rho_1\sigma_Z^2 & \rho_2\sigma_Z^2 & \cdots & 0 & 0 & 0 \\ \rho_1\sigma_Z^2 & \rho_0\sigma_Z^2 & \rho_1\sigma_Z^2 & \cdots & 0 & 0 & 0 \\ \rho_2\sigma_Z^2 & \rho_1\sigma_Z^2 & \rho_0\sigma_Z^2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \rho_0\sigma_Z^2 & \rho_1\sigma_Z^2 & \rho_2\sigma_Z^2 \\ 0 & 0 & 0 & \cdots & \rho_1\sigma_Z^2 & \rho_0\sigma_Z^2 & \rho_1\sigma_Z^2 \\ 0 & 0 & 0 & \cdots & \rho_2\sigma_Z^2 & \rho_1\sigma_Z^2 & \rho_0\sigma_Z^2 \end{pmatrix}_{n\times n},$$

where $\rho_0 = 1 + \theta_1^2 + \theta_2^2$, $\rho_1 = \theta_1(1-\theta_2)$, $\rho_2 = -\theta_2$.

For the fixed bandwidth $h_n = n^{-1/5}$ and the bandwidth based on CV, the kernel estimator $\hat{f}_n(x)$ using $k(\cdot)$ as the standard normal distribution is presented in Fig. 6. The black line in Fig. 6 is the probability density function $f(x)$ of $N(0, 0.5982)$, where 0 and 0.5982 are the estimation values of μ_Y and σ_Y^2, respectively, and the red dotted lines are the kernel estimation values. We can see that the kernel estimation has a good fitting effect in real data analysis. Overall, the fixed bandwidth and the bandwidth based on CV do not have evident differences.
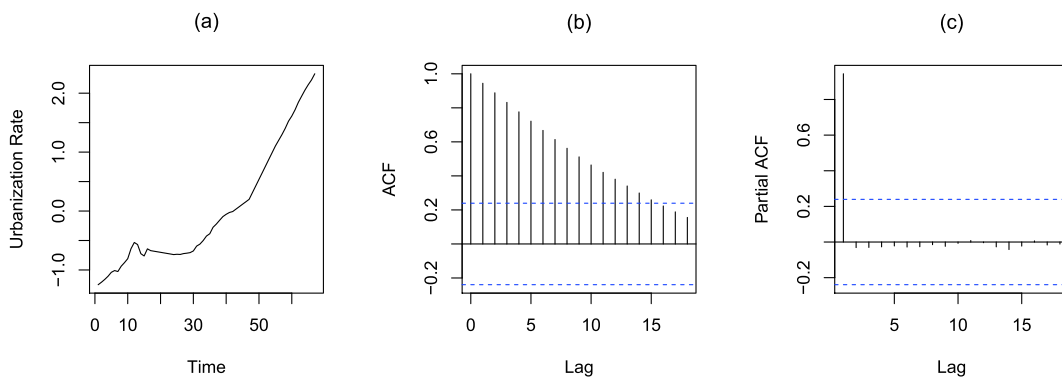
Acknowledgements

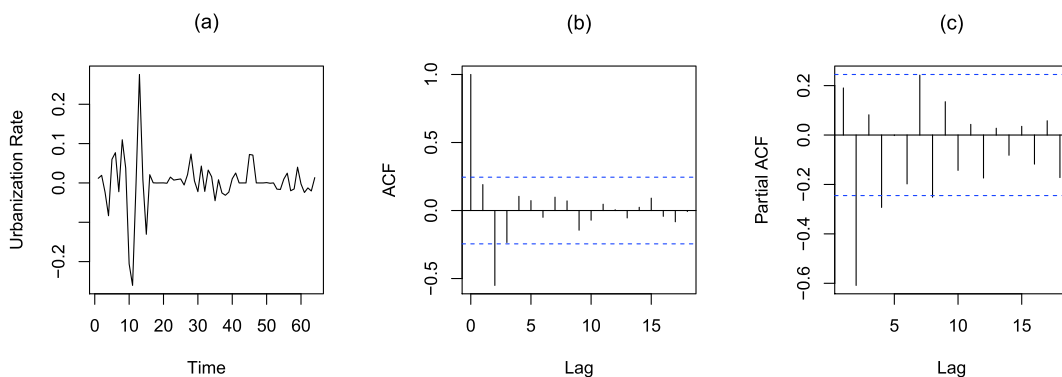**Fig. 4.** (a) Urbanization rate yearly series; (b) Autocorrelation function; (c) Partial autocorrelation function.

Fig. 5. (a) Differences of urbanization rate yearly series; (b) Autocorrelation function; (c) Partial autocorrelation function.

(a) the fixed bandwidth
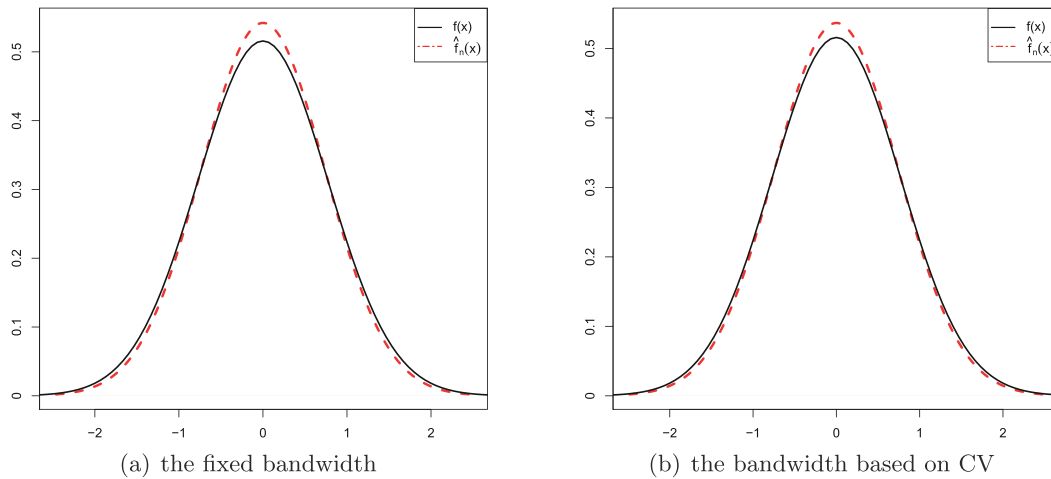

(b) the bandwidth based on CV

Fig. 6. Comparison of $\hat{f}_n(x)$ and $f(x)$ for different bandwidths.

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Wei Wang is a lecturer at School of the Big Data and Artificial Intelligence of Chizhou University. He received his Master's degree from Shanghai Maritime University in 2012. His research interests mainly focus on probability limit theory, parametric and nonparametric statistics.

Xufei Tang is a lecturer at Chaohu University. He received his Master's degree from Anhui University in 2018. His research interests focus on probability limit theory of dependent variables.

References

[1] Parzen E. On estimation of a probability density function and mode. *Ann. Math. Stat.,* **1962**, *33* (3): 1065–1076.

[2] Rosenblatt M. Curve estimates. *Ann. Math. Stat.,* **1971**, *42* (6): 1815–1842.

[3] Silverman B W. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall, 1986.

[4] Yakowitz S. Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivar. Anal.,* **1989**, *30* (1): 124–136.

[5] Bagia I, Rao B. Estimation of the survival functions for stationary associated processes. *Stat. Probab. Lett.,* **1991**, *12* (5): 291–385.

[6] Bagia I, Rao B. Kernel-type density and failure rate estimation for associated sequences. *Ann. Inst. Stat. Math.,* **1995**, *47* (2): 253–266.

[7] Roussas G G. Kernel estimates under association: Strong uniform consistency. *Stat. Probab. Lett.,* **1991**, *12* (5): 393–403.

[8] Masry E. Multivariate probability density estimation for associated processes: Strong consistency and rates. *Stat Probab. Lett.,* **2002**, *58* (2): 205–219.

[9] Oliveira P E. Density estimation for associated sampling: A point process influenced approach. *J. Nonparametr. Stat.,* **2002**, *14* (6): 495–509.

[10] Henriques C, Oliveira P E. Exponential rates for kernel density estimation under association. *Stat. Neerl.,* **2005**, *59* (4): 448–466.

[11] Ling N X, Xu C M, Peng X Z. The consistency of density function kernel estimation under NQD samples. *Journal of Hefei University of Technology (Natural Science Edition),* **2008**, *31* (2): 287–290.

[12] Su H L, Tang G Q, Sun G H. The consistency of density kernel estimation under ND samples. *Journal of Guilin University of Technology,* **2013**, *33* (3): 565–568.

[13] Wang M, Li K C. Generalized consistency of density kernel estimation. *Journal of Hubei University of Arts and Science,* **2015**, *36* (11): 19–22.

[14] Kheyri A, Amini M, Jabbari H, et al. Kernel density estimation under negative superadditive dependence and its application for real data. *Journal of Statistical Computation and Simulation,* **2019**, *89* (10): 2373–2392.

[15] Wang K Y, Wang Y B, Gao Q W. Uniform asymptotics for the finite-time ruin probability of a new dependent risk model with a constant interest rate. *Methodology and Computing in Applied Probability,* **2013**, *15*: 109–124.

[16] Joag-Dev K, Proschan F. Negative association of random variables with applications. *Annals of Statistics,* **1983**, *11* (1): 286–295.

[17] Hu T Z. Negatively superadditive dependence of random variables with applications. *Chinese Journal of Applied Probability and Statistics,* **2000**, *16* (2): 133–144.

[18] Wang X J, Xu C, Hu T C, et al. On complete convergence for widely orthant-dependent random variables and its applications in nonparametric regression models. *TEST,* **2014**, *23* (3): 607–629.

[19] Hoeffding W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.,* **1963**, *58* (301): 13–30.

[20] Liu L. Precise large deviations for dependent random variables with heavy tails. *Stat. Probab. Lett.,* **2009**, *79* (9): 1290–1298.