Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

The uniformly strong consistency of kernel-type distribution estimator under asymptotically almost negatively associated samples

Shipeng Wu, Yi Wu, Wenzhi Yang & Xuejun Wang

STATISTICAL THEORY AND RELATED FIELDS
2025, VOL. 9, NO. 2, 124–140
https://doi.org/10.1080/24754269.2025.2484980

The uniformly strong consistency of kernel-type distribution estimator under asymptotically almost negatively associated samples

Shipeng Wu[a], Yi Wu [c], Wenzhi Yang [b] and Xuejun Wang[b]

[a]School of Statistics, Shanxi University of Finance and Economics, Taiyuan, People's Republic of China; [b]School of Big Data and Statistics, Anhui University, Hefei, People's Republic of China; [c]School of Big Data and Artificial Intelligence, Chizhou University, Chizhou, People's Republic of China

ABSTRACT

This paper studies the kernel-type distribution estimator based on asymptotically almost negatively associated (AANA, for short) samples. The rate of uniformly strong consistency is established under some mild conditions. As applications, the uniformly strong convergence rates of kernel-type density estimator and kernel-type hazard rate estimator are also obtained. Some Monte Carlo simulations are presented to illustrate the finite sample performance of the kernel method. Finally, a real data analysis of Alibaba stock returns data is used to illustrate the usefulness of the proposed methodology.

ARTICLE HISTORY

Received 18 December 2023
Revised 21 January 2025
Accepted 22 March 2025

KEYWORDS

Kernel-type distribution estimator; uniformly strong consistency; AANA sequences

1. Introduction

The distribution estimation is one of the most fundamental research areas in statistical theory. Let $\{X_i\}_{i\geq 1}$ be real-valued random variables (r.v.'s, for short) having common distribution function (d.f., for short) $F(\cdot)$ and probability density function (p.d.f., for short) $f(\cdot)$ with respect to Lebesgue measure. The most common estimate of $F(\cdot)$, based on the sample $X_1,\ldots,X_n$, is the empirical d.f. $F_n(x)$ as follows

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}I(X_i\leq x),\quad x\in\mathbb{R}. \tag{1}$$

This estimate, however, does not take into consideration the smoothness of $F(\cdot)$ when the existence of a p.d.f. $f(\cdot)$ is stipulated. Hence, Yamato (1973) introduced the following kernel distribution estimator $\widehat{F}_n(\cdot)$:

$$\widehat{F}_n(x)=\frac{1}{n}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h_n}\right),\quad x\in\mathbb{R}, \tag{2}$$

where $K(\cdot)$ is a known d.f. and $h_n>0$ is called the bandwidth satisfying $h_n\to 0$ as $n\to\infty$.

Based on the assumption that $X_1,\ldots,X_n$ are independent and identically distributed (i.i.d., for short) r.v.'s., the properties of kernel distribution estimator $\widehat{F}_n(x)$ have been

CONTACT Wenzhi Yang wzyang@ahu.edu.cn School of Big Data and Statistics, Anhui University, Hefei 230601, People's Republic of China

well investigated in the literature. For example, Watson and Leadbetter (1964) showed the limit distribution of the distribution estimator $\widehat{F}_n(x)$; Yamato (1973) obtained almost uniform convergence of $\widehat{F}_n(x)$ to $F(x)$; Winter (1979) provided that $\widehat{F}_n(x)$ enjoys the Chung-Smirnov property. In addition, Azzalini (1981) established an asymptotic expression for the $\mathrm{MSE}[\widehat{F}_n(x)]$, and obtained the asymptotically optimal smoothing parameter h_n in the mean squared error (MSE, for short) sense, while Swanepoel (1988) proved that the uniform kernel is optimal in the mean integrated squared error (MISE, for short) sense and derived an expression for the bandwidth h_n.

In a number of practical application, however, the hypothesis of independence of sample $X_1, \ldots, X_n$ seems too strong, and the most suitable hypothesis is asymptotic independence. Therefore, our interest is the distribution function and its application based on dependent samples. Assuming that $X_1, \ldots, X_n$ are dependent r.v.s., there are also some studies on the properties of estimator $\widehat{F}_n(x)$ for $F(x)$. Cai and Roussas (1992) and Cai (1993) obtained almost sure uniform convergence of $\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)|$ under α-mixing and φ-mixing condition, respectively; Liu and Yang (2008) investigated the asymptotic distribution of multivariate kernel distribution estimator based on α-mixing samples. In particular, based on negatively associated (NA, for short) r.v.'s, Roussas (1995) investigated the asymptotic normality of $\sqrt{n}[\widehat{F}_n(x) - F(x)]$; Jabbari et al. (2009) studied the almost sure convergence of two-dimensional distribution function; By Lemma 4 of Yang (2003), Yang obtained the uniformly strong convergence rate of $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = O(n^{-1/2}\log^{1/2} n \log\log n)$, where $F_n(x)$ is defined by (1). Furthermore, based on asymptotically almost negatively associated random variables (AANA, for short) samples, Wu and Wang (2019) also obtained Lemma 4.5 for the uniformly strong convergence rate $F_n(x)$ as $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = O(n^{-1/2}\log^{1/2} n)$. Therefore, our works extend the empirical distribution estimator $F_n(x)$ in Yang (2003) and Wu and Wang (2019) to smooth kernel distribution estimator $\widehat{F}_n(x)$ based on AANA samples. Moreover, for more research results for empirical distribution estimator $F_n(x)$ based on other dependent case, see Lemma 2.3 in Li and Zhou (2020), Lemma 3.4 in Shen and Wang (2016), and Lemma 4.5 in Wu et al. (2022).

The NA r.v.'s were carefully studied by Joag-Dev and Proschan (1983), which pointed out a number of well-known multivariate distributions possessing the NA property. As an extension of NA r.v.'s, the AANA was introduced by Chandra and Ghosal (1996a). The concept of AANA is stated as follows.

Definition 1.1: A sequence $\{X_n, n \ge 1\}$ of random variables is called AANA if there exists a nonnegative sequence $q(n) \to 0$ as $n \to \infty$ such that

$$\begin{aligned}&\mathrm{Cov}[f(X_n),\ g(X_{n+1}, X_{n+2}, \ldots, X_{n+k})]\\ &\quad \le q(n)\{\mathrm{Var}[f(X_n)]\mathrm{Var}[g(X_{n+1}, X_{n+2}, \ldots, X_{n+k})]\}^{1/2}\end{aligned}$$

for all $n, k \ge 1$ and for all coordinatewise nondecreasing functions f and g whenever the variances exist.

Specifically, if we take $q(n) = 0$, then the concept of AANA random variables is equivalent to that of NA random variables. However, Chandra and Ghosal (1996a) also constructed an example of AANA sequences to show that some more sequences of AANA random variables are not NA. AANA is asymptotically negative correlation, while NA is a negative correlation. On the other hand, some mixing dependent sequences such as α-mixing, ρ-mixing,

φ-mixing satisfy asymptotically uncorrelated in Györfi et al. (1989). Obviously, asymptotically uncorrelated satisfies asymptotically negative correlation, but the converse is not true. The asymptotic properties related to AANA samples have been established extensively by Chandra and Ghosal (1996b), Chen et al. (2015), Shen and Wu (2014), Wang et al. (2003), Wu and Wang (2019), Yuan and An (2009, 2012) and so on.

In this paper, we mainly study the uniformly strong convergence rate of the kernel-type distribution estimator $\widehat{F}_n(x)$ based on AANA samples under some mild conditions. As applications, we further investigate the uniformly strong convergence rates for density estimator $\widehat{F}_n(x)$ and hazard rate estimator $\hat{r}_n(x)$. For the details, please see our main results in Section 3. Some simulations are performed to evaluate these kernel-type estimators in Section 4. Moreover, a real data analysis of Breast cancer data is used to illustrate the usefulness of the proposed methodology in Section 5. Finally, some conclusions are given in Section 6. The proofs of the main results are given in the Appendix. To make the R programs available to potential users, we have placed all of the codes and data sets used in this paper on the website https://github.com/proman1234/kernel-for-AANA.

2. Basic assumptions

In this section, we shall list some basic assumptions before presenting the main results.

Assumptions:

(A1) (i) (i)The sequence $\{X_i, i \geq 1\}$ is an identically distributed sequence of real-valued AANA r.v.'s with d.f. $F(x)$ and p.d.f. $f(x)$, $x \in \mathbb{R}$.
(ii) (ii)The p.d.f. $f(\cdot)$ is bounded in $\mathbb{R}$.
(iii) (iii)The second-order derivative $F''(\cdot)$ exists and it is bounded in $\mathbb{R}$.

(A2) (i) (i)$K(x)$ is a known kernel d.f. with p.d.f. $k(x)$, $x \in \mathbb{R}$.
(ii) (ii)$\int_{\mathbb{R}} uk(u)\,\mathrm{d}u = 0$ and $\int_{\mathbb{R}} u^2k(u)\,\mathrm{d}u < \infty$.
(iii) (iii)The first-order derivative $k'(\cdot)$ exists and satisfies $\int_{\mathbb{R}} |k'(u)|\,\mathrm{d}u < \infty$.

(A3) Let the bandwidth $h_n > 0$ tend to 0 as $n \to \infty$.

Remark 2.1: We list some comments on the assumptions.

(C1) The Assumptions (A1)(i)–(iii) are used commonly in the kernel distribution estimator for the random sample $\{X_n, n \geq 1\}$. For more kernel distribution estimator, see Roussas (1995), Ghorai and Susarla (1990), Liu and Yang (2008), etc.
(C2) The Assumptions (A2)(i)–(iii) are the mild conditions for kernel function $K(\cdot)$, which is available for some common kernel function such as Normal kernel, Epanechnikov kernel, Quartic and Triweight kernels, and so on. For more details, see Roussas (1995), Cai and Roussas (1999), Li et al. (2010), etc.
(C3) The Assumption (A3) is the condition for bandwidth $\{h_n, n \geq 1\}$. It is easily seen that it is a weak condition to study the properties of kernel-type distribution estimator.

3. Main results

Firstly, we investigate the uniformly strong consistency of the kernel-type estimator as follows.

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

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right| = O\left(h_n^2\right) + O\left(n^{-1/2}\log^{1/2} n\right), \quad \text{a.s.}$$

Remark 3.1: It is known that $\widehat{F}_n(x) - E\widehat{F}_n(x) = O_P(n^{-1/2})$ under some general conditions (see Li & Rcine, 2007). If we take $h_n = n^{-1/4}\log^{1/4} n$ in Theorem 3.1, then we can easily get the uniformly strong convergence rate as

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - F(x)\right| = O(n^{-1/2}\log^{1/2} n), \quad \text{a.s.} \tag{3}$$

Next, as an application of Theorem 3.1, we consider the estimator of p.d.f. $f(x)$. By taking the derivative of $\widehat{F}_n(x)$, we obtain the kernel density estimator $\widehat{F}_n(x)$ as follows:

$$\widehat{F}_n(x) = \widehat{F}_n'(x) = \frac{1}{nh_n}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h_n}\right), \tag{4}$$

where $k(\cdot) = K'(\cdot)$ is a p.d.f. of d.f. $K(\cdot)$, and h_n is a bandwidth tending to zero. In Peligrad (1992), the uniform consistency of kernel density estimator is obtained based on φ-mixing and ρ-mixing random variables. An alternative approach is to use kernel distribution that avoids the kernel function $K(\cdot)$ is monotonous.

Theorem 3.2: *Let Assumptions* (A1)–(A3) *be satisfied. If* $n^{-1}h_n^{-2}\log n \to 0$, *as* $n \to \infty$, *then,*

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - f(x)\right| = O\left(h_n\right) + O\left(n^{-1/2}h_n^{-1}\log^{1/2} n\right), \quad \text{a.s.}$$

Remark 3.2: By taking $h_n = n^{-1/4}\log^{1/4} n$, we can easily get the uniformly strong convergence rate as

$$\sup_{x\in\mathbb{R}} \left|\widehat{F}_n(x) - f(x)\right| = O\left(n^{-1/4}\log^{1/4} n\right), \quad \text{a.s.} \tag{5}$$

Based on AANA samples, Wu and Wang (2019) also studied the nearest neighbour density estimator $\tilde{f}_n(x)$ of $f(x)$, which was introduced by Loftsgarden and Quesenberry (1965) as follows:

$$\tilde{f}_n(x) = \frac{k_n}{2na_n(x)}.$$

Here, $k_n, n \geq 1$ is a sequence of positive integers such that $1 \leq k_n \leq n$ and

$$a_n(x) = \min\{a : \text{ there exist at least } k_n \text{ of } X_1, \ldots, X_n \text{ in } [x - a, x + a]\}.$$

Wu and Wang (2019) obtained the uniformly strong convergence rate $\sup_{x\in R}|\tilde{f}_n(x) - f(x)| = O(n^{-1/6}\log^{1/6} n)$ a.s., which is slightly slower than the rate $O(n^{-1/4}\log^{1/4} n)$ in (5). Therefore, our result extends the nearest neighbour density estimator $\tilde{f}_n(x)$ to smooth kernel density estimator $\widehat{F}_n(x)$ based on AANA samples.

Combining Theorem 3.1 with Theorem 3.2, we also consider the estimator of the hazard rate function, which is a basic problem in reliability theorem and biomedical science. Now, let us recall the notation of the hazard rate function $r(x)$ and its estimators $r_n(x)$ and $\hat{r}_n(x)$, which are defined as follows

$$r(x) = \frac{f(x)}{1 - F(x)}, \quad r_n(x) = \frac{f_n(x)}{1 - F_n(x)}, \quad \hat{r}_n(x) = \frac{\widehat{F}_n(x)}{1 - \widehat{F}_n(x)}, \tag{6}$$

where $F_n(x)$ and $\widehat{F}_n(x)$ are the empirical distribution estimator and kernel distribution estimator of distribution $F(x)$ with $1 - F(x) > 0$, respectively;

$$f_n(x) = \frac{1}{2nh_n} \sum_{i=1}^{n} I(x - h_n < X_i \leq x + h_n) \tag{7}$$

and $\widehat{F}_n(x)$ are the histogram density estimator (7) and the kernel density estimator (4) of $f(x)$, respectively.

Theorem 3.3: *Suppose that the conditions of Theorem* 3.2 *hold. Let* $\bar{x} = \sup\{x \in R : F(x) < 1\}$ *and the interval* $[c, d]$ *is any subset of* $(-\infty, \bar{x})$. *Then,*

$$\sup_{x \in [c,d]} \left|\hat{r}_n(x) - r(x)\right| = O(h_n) + O\left(n^{-1/2} h_n^{-1} \log^{1/2} n\right), \quad \text{a.s.}$$

Remark 3.3: Similarly to Remark 3.2, we take $h_n = n^{-1/4} \log^{1/4} n$ in Theorem 3.3 and easily get the uniformly strong convergence rate as

$$\sup_{x \in [c,d]} \left|\hat{r}_n(x) - r(x)\right| = O(n^{-1/4} \log^{1/4} n), \quad \text{a.s.} \tag{8}$$

Based on α-mixing samples, Cai and Roussas (1992) obtained the uniformly strong consistency rate as $\sup_{x \in [c,d]} |\hat{r}_n(x) - r(x)| = O(n^{-1/4} (\log \log n)^{1/4})$, which is only slightly better than the one obtained in (8). Therefore, Theorem 3.3 extends the result of Cai and Roussas (1992) from α-mixing setting to AANA setting. In addition, we give some simulations to illustrate the estimators of $\widehat{F}_n(x), F_n(x), \widehat{F}_n(x), f_n(x), \hat{r}_n(x)$ and $r_n(x)$ based on AANA samples, which show a good fit of the theoretical results.

4. Simulation

In this section, we do some simulation experiments to evaluate the finite sample performance of the kernel-type estimators in this paper. Let $\{\alpha_n, n \geq 1\}$ be a positive constant sequence with $\alpha_n \to 0$ as $n \to \infty$. The simulation data are generated from the follow scheme as

$$X_n = (1 + \alpha_n^2)^{-1/2} (Y_n + \alpha_n Y_{n+1}), \quad n \geq 1, \tag{9}$$

where $Y_n \overset{\text{i.i.d.}}{\sim} N(0, 1)$. Chandra and Ghosal (1996a) proved $\{X_n, n \geq 1\}$ to be a sequence of AANA random variables with $q(n) = O(\alpha_n)$. In addition, by the additivity of normal distribution, the distribution of $\{X_n\}_{n \geq 1}$ is identically as $N(0, 1)$, i.e. the density is $f(x) = (2\pi)^{-1/2} \exp(-x^2/2), x \in \mathbb{R}$. To compute the kernel-type estimators $\widehat{F}_n(x), \widehat{F}_n(x)$ and $\hat{r}_n(x)$, we chose the Gaussian kernel distribution function

$$K(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u, \quad x \in \mathbb{R}.$$

Consequently, the Gaussian kernel density function is taken by

$$k(x) = K'(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad x \in \mathbb{R}.$$

Moreover, we set bandwidth $h_n = [n^{-1/3}, n^{-1/4} \log^{1/4} n, n^{-1/6}]$ in our simulations. It is easily seen that the bandwidth h_n satisfies Assumption (A3). For the Cross-Validation method of selecting h_n, one can refer to Li and Rcine (2007) and the references therein.

First, for $x \in [-3, 3]$, we do some simulations for the estimators $F_n(x)$ and $\widehat{F}_n(x)$ in (1) and (2), respectively. The results are shown in Figure 1 with $n = [50, 100, 150, 200]$ for one experiment. From Figure 1, it can be seen that: (i) the fitted curves of estimator $F_n(x)$ are not as smooth as that of estimator $\widehat{F}_n(x)$; (ii) the fitted curves of $F_n(x)$ and $\widehat{F}_n(x)$ will approach the true curve of $F(x)$ as sample n increases; (iii) the different bandwidth h_n affects the fitted curves of $\widehat{F}_n(x)$, in particular for small sample n. It will be seen more clearly in Figure 2.

Next, we consider the kernel density estimator $\widehat{F}_n(x)$ defined by (4) and histogram density estimator $f_n(x)$ defined by (7), $x \in [-3, 3]$. The results are depicted in Figure 2 for one experiment. From Figure 2, we have similar conclusions as that in Figure 1 for distribution estimations. In addition, it can be seen that the fitted curves of the kernel density estimator $\widehat{F}_n(x)$ with $h_n = n^{-1/4} \log^{1/4} n$ perform better than other fitted curves.

Finally, we also compare the performance of kernel hazard estimators $\hat{r}_n$ and histogram hazard estimator $r_n(x)$ by simulation. In order to make the denominator $1 - F_n(x)$ in $r_n(x)$ away from zero in simulation, we take $x \in [-3, 3]$ to calculate hazard rate estimators $r_n(x)$ and $\hat{r}_n(x)$. Therefore, by $n = [50, 100, 200, 400]$ and $\alpha_n = n^{-2}$ in data generate model (9), the curves of estimators $r_n(x)$ and $\hat{r}_n(x)$ for $r(x)$ are presented in Figure 3 for one experiment.

Similar to Figures 2 and 3, by Figure 3, it can be seen that as sample size n increases, the kernel hazard rate estimator $\hat{r}_n(x)$ performs better than estimator $r_n(x)$, specifically $\hat{r}_n(x)$ with $h_n = n^{-1/4} \log^{1/4} n$. Thus, the kernel-type estimators $\widehat{F}_n(x)$, $\widehat{F}_n(x)$ and $\hat{r}_n(x)$ have more advantageous than the histogram estimators $F_n(x)$, $f_n(x)$ and $r_n(x)$.

Now, we do some simulations for the global error measures of asymptotic results such as Theorems 3.1–3.3. Denote the Mean Integrated Squared Error (MISE) of $\widehat{F}_n(x)$ in (2) and $F_n(x)$ in (1) with distribution function $F(x)$, respectively, as

$$\mathrm{MISE}(\widehat{F}_n) = E \int_a^b \left[\widehat{F}_n(x) - F(x)\right]^2 \mathrm{d}x, \quad [a, b] \subset \mathbb{R}, \tag{10}$$

$$\mathrm{MISE}(F_n) = E \int_a^b \left[F_n(x) - F(x)\right]^2 \mathrm{d}x, \quad [a, b] \subset \mathbb{R}. \tag{11}$$

Compared to MISE, the maximal deviation is another useful measure to compare the performances of $\widehat{F}_n(x)$ in (2) and $F_n(x)$ in (1), which are respectively defined by

$$D_n(\widehat{F}_n) = \sup_{x \in [a,b]} |\widehat{F}_n(x) - F(x)|, \quad [a, b] \subset \mathbb{R}, \tag{12}$$

$$D_n(F_n) = \sup_{x \in [a,b]} |F_n(x) - F(x)|, \quad [a, b] \subset \mathbb{R}. \tag{13}$$

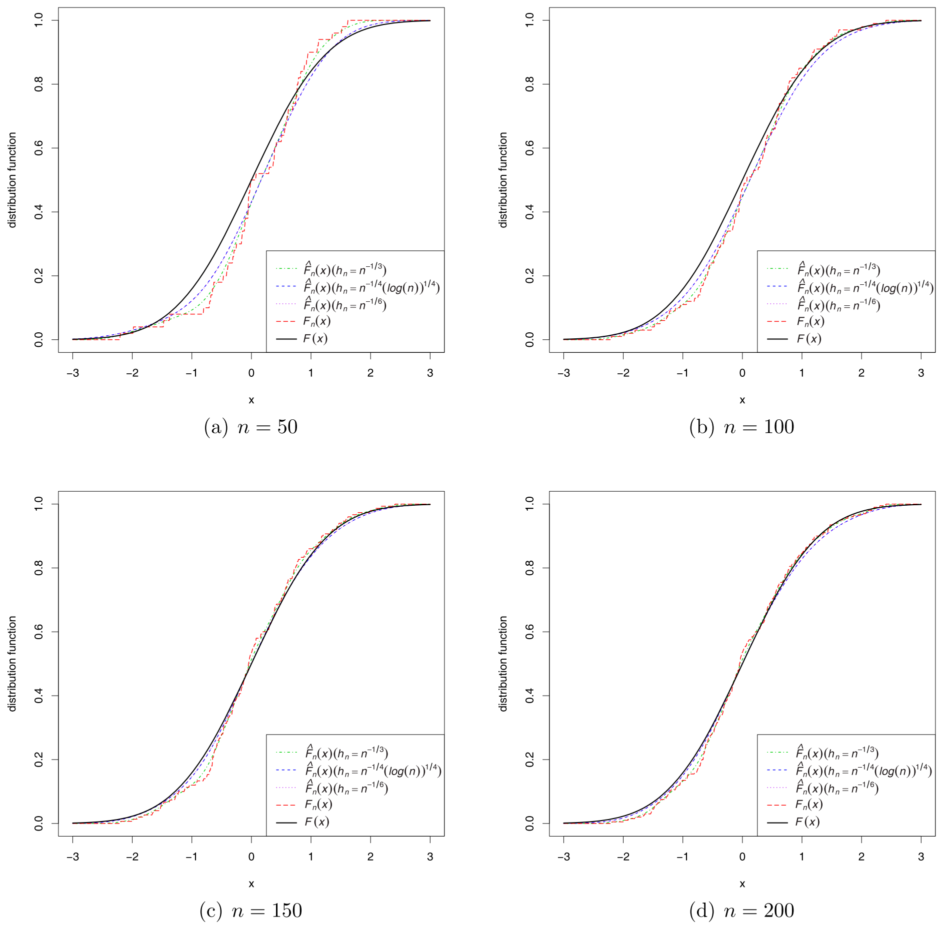


Figure 1. Comparison between $\widehat{F}_n(x)$ and $F_n(x)$ with $n = [50, 100, 150, 200]$ and $\alpha_n = n^{-2}$. (a) $n = 50$. (b) $n = 100$. (c) $n = 150$. (d) $n = 200$.

Similarly, the MISE and the maximal deviation for density estimators $\widehat{F}_n(x)$ and $f_n(x)$ and hazard rate estimators $\hat{r}_n(x)$ and $r_n(x)$ are defined as

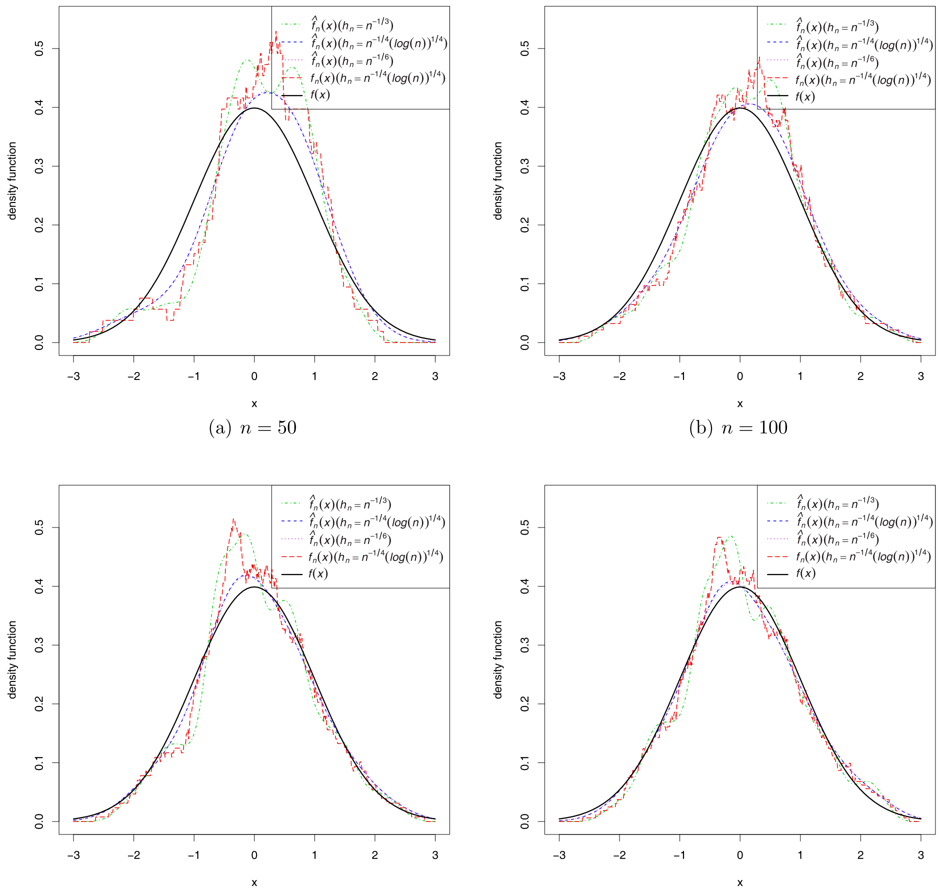
$$D_n(\widehat{F}_n) = \sup_{x\in[a,b]} |\widehat{F}_n(x) - f(x)|, \quad D_n(f_n) = \sup_{x\in[a,b]} |f_n(x) - f(x)|,$$

$$D_n(\hat{r}_n) = \sup_{x\in[c,d]} |\hat{r}_n(x) - r(x)|, \quad D_n(r_n) = \sup_{x\in[c,d]} |r_n(x) - r(x)|,$$

$$\mathrm{MISE}(\widehat{F}_n) = E\int_a^b \left[\widehat{F}_n(x) - f(x)\right]^2 \mathrm{d}x, \quad \mathrm{MISE}(f_n) = E\int_a^b \left[f_n(x) - f(x)\right]^2 \mathrm{d}x,$$

$$\mathrm{MISE}(\hat{r}_n) = E\int_c^d \left[\hat{r}_n(x) - r(x)\right]^2 \mathrm{d}x, \quad \mathrm{MISE}(r_n) = E\int_c^d \left[r_n(x) - r(x)\right]^2 \mathrm{d}x.$$

According to the above definitions, we compute the above measures with bandwidth $h_n = n^{-1/4}\log^{1/4} n$, sample sizes $n = [100, 200, 300, 400]$ and coefficients $\alpha_n = [0, n^{-1}, n^{-2}]$ in model (9). For the maximal deviations and mean integrated square errors of distribution estimators and density estimators, we take $[a, b] = [-3, 3]$ and $[c, d] = [-3, 1]$.



(a) $n = 50$ (b) $n = 100$ (c) $n = 150$ (d) $n = 200$

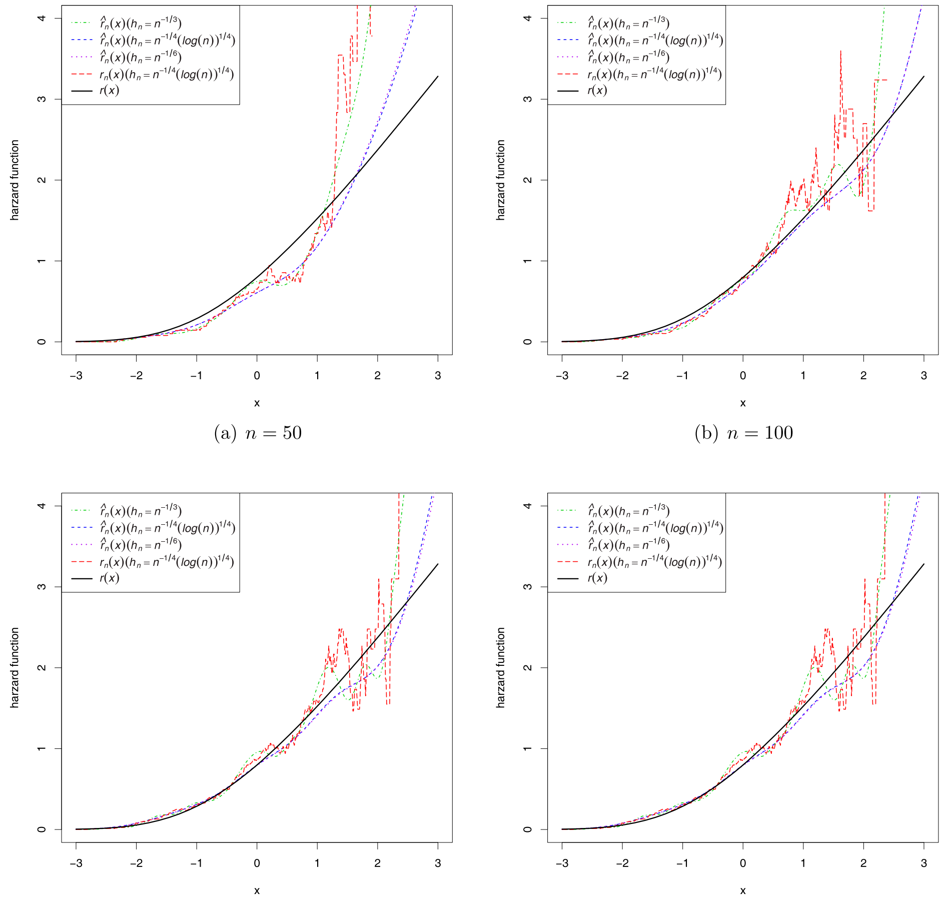
Figure 2. Comparison between $\widehat{F}_n(x)$ and $f_n(x)$ with $n = [50, 100, 150, 200]$ and $\alpha_n = n^{-2}$. (a) $n = 50$. (b) $n = 100$. (c) $n = 150$. (d) $n = 200$.

Then, we can calculate the means of $D_n(\widehat{F}_n)$, $D_n(\widehat{F}_n)$, $D_n(\hat{r}_n)$, $\text{MISE}(\widehat{F}_n)$, $\text{MISE}(\widehat{F}_n)$, $\text{MISE}(\hat{r}_n)$, by running over 1000 replications, and denote them as $\overline{D}_n(\widehat{F}_n)$, $\overline{D}_n(\widehat{F}_n)$, $\overline{D}_n(\hat{r}_n)$, $\overline{\text{MISE}}(\widehat{F}_n)$, $\overline{\text{MISE}}(\widehat{F}_n)$, $\overline{\text{MISE}}(\hat{r}_n)$, respectively. Comparing to the kernel-type estimators $\widehat{F}_n(x)$, $\widehat{F}_n(x)$ and $\hat{r}_n(x)$, we also calculate the forms

$$\frac{\overline{D}_n(F_n)}{\overline{D}_n(\widehat{F}_n)}, \quad \frac{\overline{D}_n(f_n)}{\overline{D}_n(\widehat{F}_n)}, \quad \frac{\overline{D}_n(r_n)}{\overline{D}_n(\hat{r}_n)}, \quad \frac{\overline{\text{MISE}}(F_n)}{\overline{\text{MISE}}(\widehat{F}_n)}, \quad \frac{\overline{\text{MISE}}(f_n)}{\overline{\text{MISE}}(\widehat{F}_n)}, \quad \frac{\overline{\text{MISE}}(r_n)}{\overline{\text{MISE}}(\hat{r}_n)}.$$

All the results are listed in Tables 1 and 2.

Obviously, from Tables 1 and 2, we can see that the values of $\overline{D}_n(\widehat{F}_n)$, $\overline{D}_n(\widehat{F}_n)$, $\overline{D}_n(\hat{r}_n)$, $\overline{\text{MISE}}(\widehat{F}_n)$, $\overline{\text{MISE}}(\widehat{F}_n)$ and $\overline{\text{MISE}}(\hat{r}_n)$ decrease as sample size n increases. Meanwhile, the values of $\overline{D}_n(f_n)/\overline{D}_n(\widehat{F}_n)$, $\overline{D}_n(r_n)/\overline{D}_n(\hat{r}_n)$, $\overline{\text{MISE}}(f_n)/\overline{\text{MISE}}(\widehat{F}_n)$ and $\overline{\text{MISE}}(r_n)/\overline{\text{MISE}}(\hat{r}_n)$ are bigger than 1. Only a few values of $\overline{\text{MISE}}(F_n)/\overline{\text{MISE}}(\widehat{F}_n)$ are smaller than 1, but they are very close to 1. Thus, it is concluded that the kernel-type estimators have more advantages than the histogram estimators.



(a) $n = 50$ (b) $n = 100$

(c) $n = 200$ (d) $n = 400$

Figure 3. Comparison between $\hat{r}_n(x)$ and $r_n(x)$ with $n = [50, 100, 200, 400]$ and $\alpha_n = n^{-2}$. (a) $n = 50$. (b) $n = 100$. (c) $n = 150$. (d) $n = 200$.

5. Real data analysis

The volatility is one of the characteristics of stock returns, and this feature leads to an irregular distribution. Thus, when analysing stock returns, it is worth considering the use of more flexible non-parametric estimation methods. In this section, we apply the proposed procedure to the Alibaba stock prices data sets from Jan. 3, 2023 to Dec. 31, 2024, which contain information of 501 trading days. The data can be obtained by 'getSymbols' from 'quantmod' package in R. Stock return plays a key role in several areas of finance such as asset pricing, portfolio allocation and evaluation of investment manager performance. It can be calculated by

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}, \quad t = 1, 2, \ldots, 501, \tag{14}$$

where P_t is the closing price at day t and $P_0 = P_1 \neq 0$. The daily returns on the prices of Alibaba stock are shown in Figure 4.

Table 1. Maximal deviations of distribution, density and hazard rate estimators.

		distribution estimator		density estimator		hazard rate estimator	
α_n	n	$\overline{D}_n(\widehat{F}_n)$	$\frac{\overline{D}_n(F_n)}{\overline{D}_n(\widehat{F}_n)}$	$\overline{D}_n(\widehat{F}_n)$	$\frac{\overline{D}_n(f_n)}{\overline{D}_n(\widehat{F}_n)}$	$\overline{D}_n(\hat{r}_n)$	$\frac{\overline{D}_n(r_n)}{\overline{D}_n(\hat{r}_n)}$
0	100	0.0335	1.7484	0.0424	2.0714	0.0545	6.8004
	200	0.0258	1.5934	0.0346	2.0252	0.0436	5.8751
	300	0.0231	1.5111	0.0297	2.0590	0.0405	5.2693
	400	0.0200	1.4783	0.0273	2.0605	0.0368	5.2123
$1/n$	100	0.0338	1.7603	0.0423	2.0925	0.0555	6.6576
	200	0.0262	1.6003	0.0347	2.0485	0.0454	5.9166
	300	0.0215	1.5472	0.0300	2.0579	0.0397	5.4643
	400	0.0194	1.4969	0.0268	2.0714	0.0366	5.3824
$1/n^2$	100	0.0344	1.7348	0.0422	2.1098	0.0586	6.5805
	200	0.0267	1.5872	0.0346	2.0531	0.0469	5.7562
	300	0.0222	1.5112	0.0301	2.0375	0.0409	5.6283
	400	0.0200	1.4964	0.0271	2.0516	0.0383	4.9862

Table 2. MISE of distribution, density and hazard rate estimators.

		distribution estimator		density estimator		hazard rate estimator	
α_n	n	$\overline{\text{MISE}}(\widehat{F}_n)$	$\frac{\overline{\text{MISE}}(F_n)}{\overline{\text{MISE}}(\widehat{F}_n)}$	$\overline{\text{MISE}}(\widehat{F}_n)$	$\frac{\overline{\text{MISE}}(f_n)}{\overline{\text{MISE}}(\widehat{F}_n)}$	$\overline{\text{MISE}}(\hat{r})$	$\frac{\overline{\text{MISE}}(r_n)}{\overline{\text{MISE}}(\hat{r})}$
0	100	0.0167	1.1226	0.0036	1.5310	0.0437	1.7866
	200	0.0093	1.0405	0.0022	1.5001	0.0267	1.4647
	300	0.0063	0.9872	0.0017	1.4753	0.0195	1.3834
	400	0.0049	0.9520	0.0014	1.4607	0.0161	1.3327
$1/n$	100	0.0174	1.1474	0.0036	1.5531	0.0410	1.9880
	200	0.0096	1.0259	0.0023	1.4704	0.0263	1.4724
	300	0.0066	0.9761	0.0017	1.4424	0.0201	1.3353
	400	0.0049	0.9687	0.0013	1.4714	0.0160	1.3352
$1/n^2$	100	0.0216	1.1949	0.0044	1.5354	0.0420	1.9290
	200	0.0092	1.0386	0.0022	1.4860	0.0263	1.5297
	300	0.0063	0.9888	0.0017	1.4892	0.0201	1.3679
	400	0.0048	0.9574	0.0013	1.4698	0.0160	1.3000

From Figure 4, it can be seen that the time series of return $\{R_t, 1 \le t \le 501\}$ is stationary. So we will apply the kernel-type estimator and non-smooth estimator to fit the distribution of this data. Similar to Li et al. (2023), the picture of sample autocorrelation functions (ACF) with samples $R_1, R_2, \ldots, R_{501}$ is presented in Figure 5. It can be seen that the sequence $\{R_1, R_2, \ldots, R_{501}\}$ is asymptotically uncorrelated. Thus, we assume it to be AANA in this real data analysis.

Next, we consider distribution estimators with data $\{R_t, 1 \le t \le 501\}$. The kernel distribution estimator $\widehat{F}_n(x)$ with bandwidth $h_n = n^{-1/4}\log^{1/4} n$ and empirical distribution estimator $F_n(x)$, respectively, are shown in the Figure 6 (left), where the solid black line is a normal distribution function with mean $\bar{R} = \sum_{t=1}^{501} R_t/501 = 0.0001264598$ and variance $s^2(R) = \sum_{t=1}^{501}(R_t - \bar{R})^2/500 = 0.0006178045$, denoted by $F(x)$. It is pointed out that this normal distribution function $F(x)$ is not the true distribution of data $\{R_t, 1 \le t \le 501\}$.

Although these two distribution estimators $\widehat{F}_n(x)$ and $F_n(x)$ are in good agreement, it is more important to make sure that the same is true for the density functions, because the closeness of two distribution function estimators does not imply the closeness of the corresponding density functions. The Figure 6 (right) displays that the density function obtained by using the two density estimator $\widehat{F}_n(\cdot)$ and $f_n(\cdot)$. It is seen from this figure that the shapes of

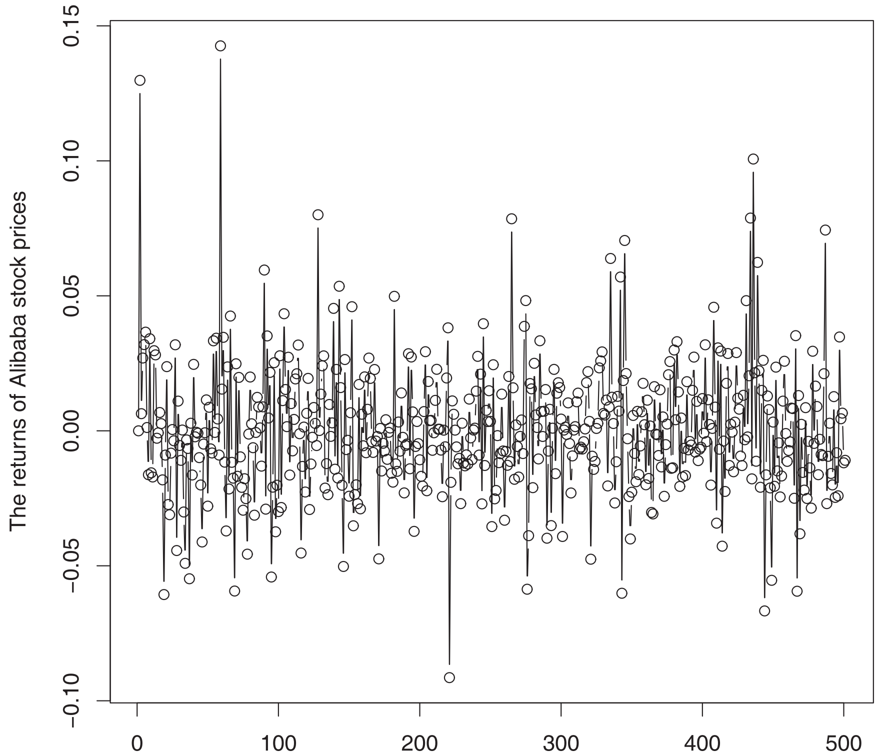


Figure 4. Times series of returns of Alibaba stock from Jan. 3, 2023 to Dec. 31, 2024.

the density estimators $\widehat{F}_n(\cdot)$ and $f_n(\cdot)$ are significantly different with the normal density function $f(x)$ of $F(x)$. We can observe that the distribution of Alibaba stock returns is leptokurtic, which is consistent with the findings from Mikael (2001). Therefore, the non-parametric kernel estimation method not only has more advantages than ordinary normal distribution in capturing the market risk characteristics, but also can better reflect the characteristics of peak of the stock returns distribution. At the same time, the conclusion is more credible.

6. Conclusion

In this paper, we have considered the kernel-type estimators for the distribution function, density function and hazard rate function based on the AANA samples, and obtained the uniformly strong convergence rates for these estimators. In practice, we can test whether a time series is an AANA by computing the sample autocorrelation function, since AANA is asymptotically negative correlation. Our simulation studies indicate that the kernel-type estimators have good performance compared to the histogram estimators. In particular, we advise to use kernel estimator $\hat{r}_n(x)$ to estimate hazard function rather than $r_n(x)$. In the real data analysis, it will be seen that the non-parametric kernel estimation method is particularly suitable for estimating the distribution function of stock returns. It is known that AANA sequence is asymptotically negative correlation, while the mixing sequences are asymptotically uncorrelated. Many works of estimators $\widehat{F}_n(x)$, $\widehat{F}_n(x)$ and $\hat{r}_n(x)$ with mixing conditions can be found in Györfi et al. (1989), and it is interesting for research to study them with other

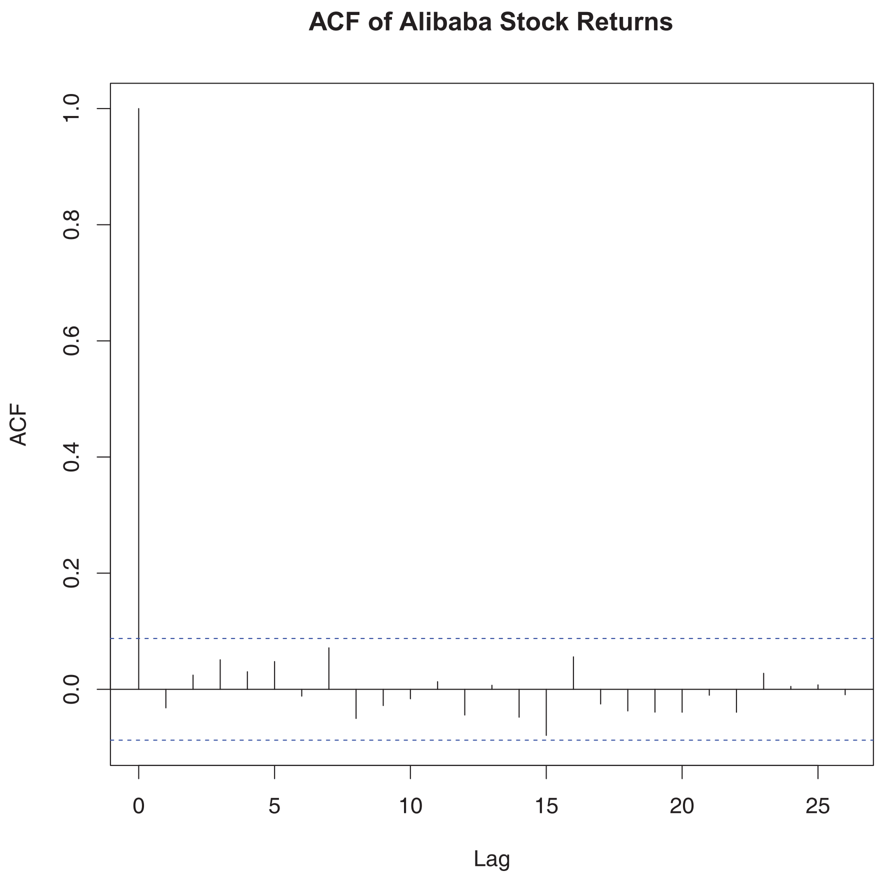


Figure 5. Sample autocorrelation functions (ACF) plots for Alibaba returns data series.

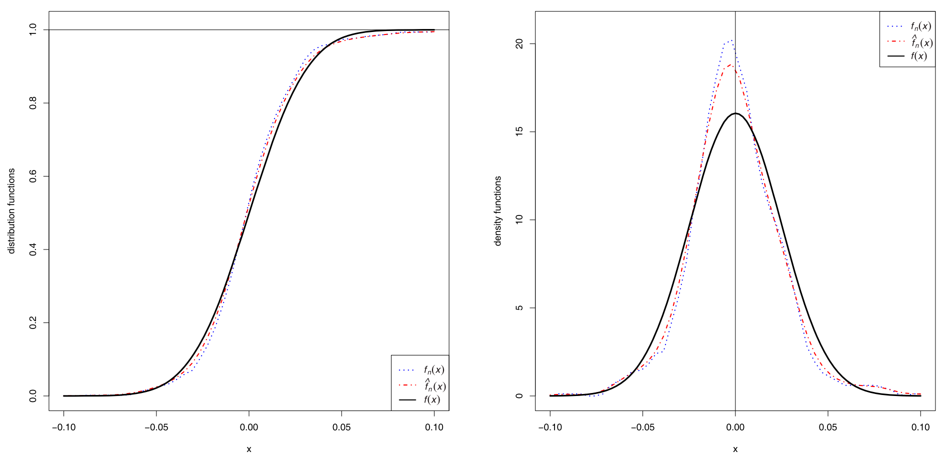


Figure 6. Distribution function (left) and density function (right) for Alibaba stock returns, where the blue lines base on histogram estimators, and the black lines are normal distribution function curves. (a) $n = 50$. (b) $n = 100$.

dependent samples. In this paper, we did not obtain the optimal uniformly convergence rates with AANA samples, compared to the independent samples. Thus, we will pay attention to the study of optimal uniformly convergence rate, asymptotic distribution and the applications in future work.

Acknowledgments

The authors would like to thank the editor and two anonymous referees for their insightful comments and suggestions, which improved the manuscript very much.

Disclosure statement

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

Funding

This work is supported by National Natural Science Foundation of China [grant number 11671012], Natural Science Foundation of Anhui Province [grant numbers 2008085MA14 and 2108085MA06], Scientific and Technological Innovation Project of Higher Education Institutions in Shanxi Province [grant number 23L165] and General Projects under the 2023 Annual Research Plan of the China Society for Business Statistics [grant number 2023STY62].

ORCID

Yi Wu http://orcid.org/0000-0003-2635-052X
Wenzhi Yang http://orcid.org/0000-0003-0487-5974

References

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, *68*(1), 326–328.

Cai, Z. (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*(1), 11–17.

Cai, Z., & Roussas, G. G. (1992). Uniform strong estimation under α-mixing, with rates. *Statistics & Probability Letters*, *15*(1), 47–55.

Cai, Z., & Roussas, G. G. (1999). Berry-Esseen bounds for smooth estimator of a distribution function under association. *Journal of Nonparametric Statistics*, *11*(1-3), 79–106.

Chandra, T. K., & Ghosal, S. (1996a). Extensions of the strong law of large numbers of Marcinkiewicz and Zygmund for dependent variables. *Acta Mathematica Hungarica*, *71*(4), 327–336.

Chandra, T. K., & Ghosal, S. (1996b). The strong law of large numbers for weighted averages under dependence assumptions. *Journal of Theoretical Probability*, *9*(3), 797–809. https://doi.org/10.1007/BF02214087

Chen, Z. Y., Wang, X. J., & Hu, S. H. (2015). Strong laws of large numbers for weighted sums of asymptotically almost negatively associated random variables. *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas. RACSAM*, *109*(1), 135–152.

Ghorai, J. K., & Susarla, V. (1990). Kernel estimation of a smooth distribution function based on censored data. *Metrika*, *37*(2), 71–86. https://doi.org/10.1007/BF02613509

Györfi, L., Härdle, W., Sarda, P., & Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series*. Springer.

Jabbari, H., Azarnoosh, H. A., & Fakoor, V. (2009). Almost sure convergence of two-dimensional distribution function under negative association. *Journal of Applied Probability and Statistics*, *4*(2), 157–166.

Joag-Dev, K., & Proschan, F. (1983). Negative association of random variables with applications. *The Annals of Statistics*, *11*(1), 286–295. https://doi.org/10.1214/aos/1176346079
Li, Q., & Rcine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Li, Y. M., Wei, C., & Yang, S. C. (2010). The recursive kernel distribution function estimator based on negatively and positively associated sequences. *Communications in Statistics–Theory and Methods*, *39*(20), 3585–3595. https://doi.org/10.1080/03610920903427750
Li, X. Q., Zhang, L., Shen, Y., & Chen, Z. Y. (2023). Asymptotic approximations of random ratio model based on AANA sequences. *Communications in Statistics-Simulation and Computation*, *52*(8), 3796–3819. https://doi.org/10.1080/03610918.2021.1948573
Li, Y. M., & Zhou, Y. (2020). The Kaplan-Meier estimator and hazard estimator for censored END survival time observations. *Communications in Statistics–Theory and Methods*, *49*(11), 2690–2702. https://doi.org/10.1080/03610926.2019.1580737
Liu, R., & Yang, L. (2008). Kernel estimation of multivariate cumulative distribution function. *Journal of Nonparametric Statistics*, *20*(8), 661–677. https://doi.org/10.1080/10485250802326391
Loftsgarden, D. O., & Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, *36*(3), 1049–1051. https://doi.org/10.1214/aoms/1177700079
Mikael, L. (2001). A model for stock return distribution. *International Journal of Finance & Economics*, *6*(2), 159–169. https://doi.org/10.1002/ijfe.v6:2
Peligrad, M. (1992). Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumptions. *Stochastics and Stochastic Reports*, *40*, 147–168. https://doi.org/10.1080/17442509208833786
Roussas, G. G. (1995). Asymptotic normality of a smooth estimate of a random field distribution function under association. *Statistics & Probability Letters*, *24*(1), 77–90. https://doi.org/10.1016/0167-7152(94)00151-W
Shen, A. T., & Wang, X. H. (2016). Kaplan-Meier estimator and hazard estimator for censored negatively superadditive dependent data. *Statistics*, *50*(2), 377–388.
Shen, A. T., & Wu, R. C. (2014). Strong convergence for sequences of asymptotically almost negatively associated random variables. *Stochastics*, *86*(2), 291–303. https://doi.org/10.1080/17442508.2013.775289
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*(11), 3785–3799. https://doi.org/10.1080/03610928808829835
Wang, Y. B., Yan, J. G., & Cheng, F. Y. (2003). The strong law of large numbers and the iterated logarithm for product sums of NA and AANA random variables. *Southeast Asian Bulletin of Mathematics*, *27*(2), 369–384.
Watson, G. S., & Leadbetter, M. R. (1964). Hazard analysis II. *Sankhyā: The Indian Journal of Statistics, Series A*, *26*, 101–116.
Winter, B. B. (1979). Convergence rate of perturbed empirical distribution functions. *Journal of Applied Probability*, *16*(1), 163–173. https://doi.org/10.2307/3213384
Wu, Y., & Wang, X. J. (2019). On consistency of the nearest neighbor estimator of the density function and its applications. *Acta Mathematica Sinica (English Series)*, *35*(5), 703–720. https://doi.org/10.1007/s10114-019-8099-9
Wu, Y., Yu, W., & Wang, X. J. (2022). Strong representations of the Kaplan-Meier estimator and hazard estimator with censored widely orthant dependent data. *Computational Statistics*, *37*(1), 383–402. https://doi.org/10.1007/s00180-021-01125-z
Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, *15*(3–4), 69–78. https://doi.org/10.5109/13073
Yang, S. C. (2003). Consistency of nearest neighbor estimator of density function for negative associated samples. *Acta Mathematicae Applicatae Sinica (English Series)*, *26*(3), 385–394.
Yuan, D., & An, J. (2009). Rosenthal type inequalities for asymptotically almost negatively associated random variables and applications. *Science in China Series A: Mathematics*, *52*(9), 1887–1904. https://doi.org/10.1007/s11425-009-0154-z

Yuan, D., & An, J. (2012). Laws of large numbers for Cesàro alpha-integrable random variables under dependence condition AANA or AQSI. *Acta Mathematica Sinica (English Series)*, *28*(6), 1103–1118. https://doi.org/10.1007/s10114-012-0033-3

Appendix. Proofs of Theorems 3.1, 3.2 and 3.3

Lemma A.1 (Wu and Wang (2019), Lemma 4.5.): *Let $\{X_n, n \geq 1\}$ be a sequence of AANA random variables with continuous distribution function $F(x)$, $x \in \mathbb{R}$. Then*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = O\left(n^{-1/2}\log^{1/2} n\right) \quad \text{a.s.,}$$

where $F_n(x)$ is defined by (1).

A.1 Proof of Theorem 3.1

According to the definition of $\widehat{F}_n(x)$ in (2) and Assumption (A2)(i), we have

$$\begin{aligned}\widehat{F}_n(x) &= \int_{-\infty}^{+\infty} K\left(\frac{x-y}{h_n}\right) \mathrm{d}F_n(y) \\ &= K\left(\frac{x-y}{h_n}\right) F_n(y)\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} F_n(y)\,\mathrm{d}K\left(\frac{x-y}{h_n}\right) \\ &= \int_{-\infty}^{+\infty} F_n(x-h_nu)k(u)\,\mathrm{d}u,\end{aligned}$$

where $F_n(x)$ is defined by (1). Furthermore, it follows that

$$\begin{aligned}\widehat{F}_n(x) - F(x) &= \int_{-\infty}^{+\infty} [F_n(x-h_nu) - F(x-h_nu)]\,k(u)\,\mathrm{d}u \\ &\quad + \int_{-\infty}^{+\infty} [F(x-h_nu) - F(x)]k(u)\,\mathrm{d}u. \end{aligned} \tag{A1}$$

By Lemma A.1 and Assumption (A2)(i), it is easy to see that

$$\sup_x \left| \int_{-\infty}^{+\infty} [F_n(x-h_nu) - F(x-h_nu)]k(u)\,\mathrm{d}u \right| = O(n^{-1/2}\log^{1/2} n) \quad \text{a.s.} \tag{A2}$$

By Taylor's expansion to $F(x)$, we have

$$F(x-h_nu) - F(x) = -F'(x)h_nu + \frac{F''(\xi)}{2}h_n^2u^2, \tag{A3}$$

where ξ lies between x and $x - h_nu$. Then, combining this with Assumptions (A1)(ii)-(iii) and (A2)(ii), we obtain

$$\begin{aligned}\int_{-\infty}^{+\infty} [F(x-h_nu) - F(x)]k(u)\,\mathrm{d}u &= -h_nF'(x)\int_{-\infty}^{\infty} uk(u)\,\mathrm{d}u + \frac{h_n^2F''(\xi)}{2}\int_{-\infty}^{\infty} u^2k(u)\,\mathrm{d}u \\ &= O\left(h_n^2\right).\end{aligned}$$

This implies

$$\sup_x \left| \int_{-\infty}^{+\infty} [F(x-h_nu) - F(x)]\,k(u)\,\mathrm{d}u \right| = O\left(h_n^2\right). \tag{A4}$$

Combining (A1)–(A4), we obtain

$$\sup_x \left|\widehat{F}_n(x) - F(x)\right| = O\left(n^{-1/2}\log^{1/2} n\right) + O\left(h_n^2\right) \text{ a.s.} \tag{A5}$$

So, we complete the proof of Theorem 3.1.

A.2 Proof of Theorem 3.2

Applying the integrating by parts and change of variable, we obtain

$$
\begin{aligned}
\widehat{F}_n(x) - f(x) &= \frac{1}{nh_n}\sum_{i=1}^{n} k\left(\frac{x-X_i}{h_n}\right) - f(x) \\
&= \frac{1}{h_n}\int_{-\infty}^{+\infty} k\left(\frac{x-t}{h_n}\right) \mathrm{d}F_n(t) - \frac{1}{h_n}\int_{-\infty}^{+\infty} k\left(\frac{x-t}{h_n}\right) \mathrm{d}F(t) \\
&\quad + \frac{1}{h_n}\int_{-\infty}^{+\infty} k\left(\frac{x-t}{h_n}\right) \mathrm{d}F(t) - f(x)\int_{-\infty}^{+\infty} k\,(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{+\infty} k\,(u)\,\mathrm{d}F_n(x-h_nu) - \int_{-\infty}^{+\infty} k\,(u)\,\mathrm{d}F(x-h_nu) \\
&\quad + \int_{-\infty}^{+\infty} k\,(u)\,\mathrm{d}F(x-h_nu) - f(x)\int_{-\infty}^{+\infty} k\,(u)\,\mathrm{d}u \\
&= k(u)F_n(x-h_nu)|_{-\infty}^{\infty} - k(u)F(x-h_nu)|_{-\infty}^{\infty} \\
&\quad - \int_{-\infty}^{+\infty} [F_n(x-h_nu) - F(x-h_nu)]\,\mathrm{d}k\,(u) \\
&\quad + \int_{-\infty}^{+\infty} k\,(u)\,[f(x-h_nu) - f(x)]\,\mathrm{d}u. \qquad \text{(A6)}
\end{aligned}
$$

By $\int_{-\infty}^{\infty} u^2k(u)\,\mathrm{d}u < \infty$ in Assumption (A2)(ii), we have $k(u) \to 0$ as $|u| \to \infty$. Thus, it has $k(u)F_n(x-h_nu)|_{-\infty}^{\infty} = 0$ and $k(u)F(x-h_nu)|_{-\infty}^{\infty} = 0$.

Next, by Theorem 3.1 and $\int_{-\infty}^{\infty} |k'(u)|\,\mathrm{d}u < \infty$ in Assumption (A2)(iii), we have

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|\int_{-\infty}^{+\infty} F_n(x-h_nt) - F(x-h_nt)\,\mathrm{d}k\,(t)\right| \\
&\le \sup_{(x-h_nu)\in\mathbb{R}} |F_n(x-h_nu) - F(x-h_nu)| \int_{-\infty}^{+\infty} \left|k'\,(u)\right|\mathrm{d}u \\
&= O(h_n^2) + O\left(n^{-1/2}h_n^{-1}\log^{1/2} n\right) \quad \text{a.s.} \qquad \text{(A7)}
\end{aligned}
$$

By $\int_{-\infty}^{\infty} uk(u)\,\mathrm{d}u = 0$ in Assumption (A2)(ii), it has $\int_{-\infty}^{\infty} |u|k(u)\,\mathrm{d}u < \infty$. Combining this and the boundness of $f'(x)$ in Assumption (A1)(iii), we obtain that

$$
\int_{-\infty}^{+\infty} k\,(u)\left|f(x-h_nu) - f(x)\right|\mathrm{d}u \le h_n|f'(\eta)|\int_{-\infty}^{+\infty} k\,(u)\,|u|\,\mathrm{d}u = O(h_n), \qquad \text{(A8)}
$$

where η lies between x and $x - h_nu$.

Therefore, by (A6)–(A8), it has

$$
\sup_{x\in\mathbb{R}}\left|\widehat{F}_n(x) - f(x)\right| = O(h_n) + O(n^{-1/2}h_n^{-1}\log^{1/2} n) \quad \text{a.s.}
$$

Hence, we complete the proof of Theorem 3.2.

A.3 Proof of Theorem 3.3

By elementary calculus, we have

$$
\hat{r}_n(x) - r(x) = \frac{[1-F(x)][\widehat{F}_n(x) - f(x)] + f(x)[\widehat{F}_n(x) - F(x)]}{[1-F(x)][1-\widehat{F}_n(x)]}. \qquad \text{(A9)}
$$

By Theorem 3.1, we know that $\sup_{x\in[c,d]} |\widehat{F}_n(x) - F(x)| \to 0$ a.s. Thus, $(1-F(x))(1-\widehat{F}_n(x))$ is bounded away from zero uniformly in $x \in [c,d]$. Then, utilizing Theorems 3.1 and 3.2 and Assumption

(A1)(ii), we get

$$\sup_{x\in[c,d]} |\hat{r}_n(x) - r(x)| = O(h_n) + O(n^{-1/2}h_n^{-1}\log^{1/2} n) \quad \text{a.s.}$$

Thus, we complete the proof of Theorem 3.3.