COMPLETE CONSISTENCY AND CONVERGENCE RATE OF THE NEAREST NEIGHBOR ESTIMATOR OF THE DENSITY FUNCTION BASED ON WOD SAMPLES

KAN CHEN*, XUEFEI TANG AND JIANGFENG HAO

(Communicated by X. Wang)

Abstract. By using the exponential inequality of widely orthant dependent (WOD, for short) random variables, we mainly investigate the complete consistency and convergence rate of the nearest neighbor estimator of the density function based on WOD samples. The results obtained in the paper generalize and improve some corresponding ones in the literature. In addition, the restriction on the dominating coefficients \( g(n) \) is much weak, even if the geometric growth of \( g(n) \), the consistency result and convergence rate still hold by using the results that we obtained.

1. Introduction

The estimation of probability density function has important applications in medicine, engineering and economy. Therefore, the estimation of probability density function is still a hot research topic. There are many methods for its estimation, such as kernel estimation, wavelet estimation, maximum likelihood estimation, and so on. Suppose that the population \( X \) has an unknown density function \( f(x) \), \( X_1, X_2, \ldots, X_n \) are the samples from the population \( X \). Let \( \{k_n, n \geq 1\} \) be a sequence of positive integers, such that \( 1 \leq k_n \leq n \). For fixed \( x \) and \( n \), denote

\[
a_n(x) = \min\{a : \text{there exist at least } k'_ni \text{ such that } X_i \in [x-a, x+a]\}.
\]

Loftsgarden and Quesenberry [1] proposed the following nearest neighbor estimator of the density function \( f(x) \):

\[
f_n(x) = \frac{k_n}{2na_n(x)}. \tag{1.1}
\]

Since Loftsgarden and Quesenberry [1] put forward the method of nearest neighbor estimation of density function mentioned above, many scholars have studied the


Keywords and phrases: WOD samples, nearest neighbor estimator, complete consistency, convergence rate.

Supported by the Key Research Project of Chaohu University (XLZ-201903, XLZ-201904, XLY-201905), the General Projects for Supporting Outstanding Young Talents in Universities in Anhui Province (gxyq2019082) and the Key Projects of Natural Science Research in Universities in Anhui Province (KJ2019A0683; KJ2020A0679).

* Corresponding author.
asymptotic properties for this estimator. Based on independent samples, Loftsgarden and Quesenberry [1] established the weak consistency; Wagner [2] studied the strong consistency; Moore and Henrichon [3], and Devroye and Wagner [4] obtained the uniform consistency and the uniformly strong consistency, respectively; Chen [5] derived the convergence rate of the consistency, and so on. Under dependent samples, there are also many interesting results obtained by many scholars. For example, Boente and Fraiman [6] established the strong consistency for the nearest neighbor estimator based on \( \varphi \)-mixing and \( \alpha \)-mixing samples; Chai [7] obtained the weak consistency, strong consistency, uniformly strong consistency and the convergence rate under stationary \( \varphi \)-mixing samples; Liu and Zhang [8] established the asymptotic normality for the nearest neighbor estimator based on \( \varphi \)-mixing samples; Yang [9] investigated the weak consistency, strong consistency, uniformly strong consistency and the corresponding convergence rate under negatively associated (NA, for short) samples; Wang and Hu [10] extended the result of Yang [9] from NA samples to widely orthant dependent (WOD, for short) samples.

In this paper, we will continue to study the strong consistency, uniformly strong consistency and the corresponding convergence rate for the nearest neighbor density estimator based on WOD samples. Let’s first review the concept of WOD random variables proposed by Wang et al. [11] as follows:

**Definition 1.1.** A sequence \( \{X_n; n \geq 1\} \) of random variables is said to be widely orthant dependent (WOD, for short), if there exist two positive sequences \( \{g_U(n), n \geq 1\} \) and \( \{g_L(n), n \geq 1\} \), such that for each \( n \geq 1 \) and all real numbers \( x_1, x_2, \ldots, x_n \), both

\[
P(X_1 > x_1, X_2 > x_2, \ldots, X_n > x_n) \leq g_U(n) \prod_{i=1}^{n} P(X_i > x_i)
\]

and

\[
P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) \leq g_L(n) \prod_{i=1}^{n} P(X_i \leq x_i)
\]

hold.

An array \( \{X_{ni}; i \geq 1, n \geq 1\} \) of random variables is said to be rowwise WOD, if for each \( n \geq 1 \), \( \{X_{ni}; i \geq 1\} \) is WOD. \( g_U(n), g_L(n), n \geq 1 \) are called dominating coefficients.

Denote \( g(n) = \max\{g_U(n), g_L(n)\} \). It is easy to check that \( g_U(n) \geq 1 \) and \( g_L(n) \geq 1 \), and thus \( g(n) \geq 1 \). If \( g_U(n) = g_L(n) = M \) for all \( n \geq 1 \), where \( M \geq 1 \) is a positive constant, then WOD random variables are degenerated to extended negatively dependent (END, for short) random variables, which were introduced by Liu [12] in the year 2009; If \( g_U(n) = g_L(n) = 1 \), then WOD random variables are degenerated to negatively orthant dependent (NOD, for short) random variables, which were introduced by Lehmann [13] in the year 1966, and carefully studied by Joag-Dev and Proschan [14]. It is well known that NA random variables are NOD, and thus are WOD. Hu [15] pointed out negatively superadditive dependent (NSD, for short) random variables are NOD, and thus are WOD. So, WOD is a kind of very broad dependent structure which includes NA random variables, NSD random variables, NOD random variables, END
random variables, and some positive dependent structures. The study of its limit properties is also of great theoretical and practical significance. For more details about the WOD random variables, one can refer to [16]–[25].

In this work, we will use the exponential inequality of WOD random variables to study the complete consistency, uniformly complete consistency and the corresponding convergence rate of the nearest neighbor density estimator under the WOD samples, and the condition on the dominating coefficients $g(n)$ is very common. In addition, the convergence rate obtained in the paper also improves the corresponding ones of Yang [9] and Wang and Hu [10]. Throughout the paper, $[x]$ stands for the integer part of $x$, $C$ and $c_0$ represent positive constants whose values may vary in different places. Denote $\log x = \ln \max(x, e)$, where $\ln$ represents the natural logarithm. Let $I(A)$ be the indicator function of the set $A$. $c(f)$ denotes all the continuity points of the function $f$, and $X_n \rightarrow C$ a.c. stands for $\sum_{n=1}^{\infty} P(|X_n - C| > \varepsilon) < \infty$ for any $\varepsilon > 0$, that is, the sequence $\{X_n, n \geq 1\}$ of random variables converges completely to $C$.

2. Preliminary lemmas

To prove the main results of the paper, we need the following important lemmas. The first one is a basic property for WOD random variables, which can be found in Wang et al. [21].

**Lemma 2.1.** Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables. If $\{f_n(\cdot), n \geq 1\}$ are all nondecreasing (or all nonincreasing), then $\{f_n(X_n), n \geq 1\}$ are still WOD.

The next one is the Bernstein type inequality for WOD random variables, which has been proved by Xia et al. [26].

**Lemma 2.2.** Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables with mean zero and $|X_n| \leq d_n$ a.s. for each $n \geq 1$, where $\{d_n, n \geq 1\}$ is a sequence of positive numbers. Denote $b_n = \max_{1 \leq i \leq n} d_i$ and $\Delta_n^2 = \sum_{i=1}^{n} EX_i^2$ for each $n \geq 1$. Then for any $\varepsilon > 0$,

$$P\left(\left| \sum_{i=1}^{n} X_i \right| \geq \varepsilon \right) \leq 2g(n) \exp\left\{-\frac{\varepsilon^2}{2(2\Delta_n^2 + b_n\varepsilon)}\right\}.$$ 

**Lemma 2.3.** (cf. Yang [9]) Let $F(x)$ be a continuous distribution function. For $n \geq 3$, assume that $x_{nj}$’s satisfy $F(x_{nj}) = j/n, j = 1, 2, \ldots, n-1$. Then

$$\sup_{-\infty < x < \infty} |F_n(x) - F(x)| \leq \max_{1 \leq j \leq n-1} |F_n(x_{nj}) - F(x_{nj})| + 2/n, \quad (2.1)$$

where $F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i < x)$ is the empirical distribution function of $X_1, X_2, \ldots, X_n$. 

LEMMA 2.4. Let \( \{X_n, n \geq 1\} \) be a sequence of WOD random variables with the dominating coefficients \( g(n) \), unknown distribution function \( F(x) \) and bounded density function \( f(x) \). Let \( F_n(x) \) be the empirical distribution function of \( X_1, X_2, \cdots, X_n \), and \( \{\kappa_n, n \geq 1\} \) be a sequence of positive numbers such that \( \kappa_n \to 0 \) and \( \liminf_{n \to \infty} n\kappa_n^2/\log(ng(n)) \geq c_0 > 0 \). Then for any sufficiently large \( D_0 > 0 \),

\[
\sum_{n=1}^{\infty} P \left( \sup_x |F_n(x) - F(x)| > D_0 \kappa_n \right) < \infty.
\]

In particularly, we have

\[
\sum_{n=1}^{\infty} P \left( \sup_x |F_n(x) - F(x)| > D_0 (\log(ng(n))/n)^{1/2} \right) < \infty.
\]

Proof. Noting that \( n\kappa_n \to \infty \), we have \( 2/n < D_0 \kappa_n / 2 \) for all sufficient large \( n \) and any positive constant \( D_0 \), which together with (2.1) yields that

\[
P \left( \sup_x |F_n(x) - F(x)| > D_0 \kappa_n \right) \leq P \left( \max_{1 \leq j \leq n-1} |F_n(x_{nj}) - F(x_{nj})| > D_0 \kappa_n / 2 \right)
\]

\[
\leq \sum_{j=1}^{n-1} P \left( |F_n(x_{nj}) - F(x_{nj})| > D_0 \kappa_n / 2 \right). \tag{2.2}
\]

Denote \( X_i(x_{nj}) = I(X_i < x_{nj}) - EI(X_i < x_{nj}) \). It follows by Lemma 2.1 that \( \{X_i(x_{nj}), i \geq 1\} \) is still a sequence of WOD random variables with \( EX_i(x_{nj}) = 0 \), \( |X_i(x_{nj})| \leq 1 \) and \( E(X_i(x_{nj}))^2 \leq 1 \). Hence, we have by Lemma 2.2 that for all sufficiently large \( n \),

\[
P(|F_n(x_{nj}) - F(x_{nj})| > D_0 \kappa_n / 2) = P \left( \left| \sum_{i=1}^{n} X_i(x_{nj}) \right| > D_0 n \kappa_n / 2 \right)
\]

\[
\leq C g(n) \exp \left\{ - \frac{D_0^2 n^2 \kappa_n^2}{16 \Delta_n^2 + 4D_0 n \kappa_n} \right\}
\]

\[
\leq C g(n) \exp \left\{ - \frac{D_0^2}{18} n \kappa_n^2 \right\}
\]

\[
\leq C g(n) \exp \left\{ - \frac{c_0 D_0^2}{18} \log(ng(n)) \right\}
\]

\[
\leq C g(n)^{1-c_0 D_0^2/18} n^{-c_0 D_0^2/18}. \tag{2.3}
\]

Take \( D_0 \) sufficiently large such \( c_0 D_0^2 / 18 > 2 \). It follows by \( g(n) \geq 1 \) that \( g(n)^{1-c_0 D_0^2/18} \leq 1 \). Thus, we have by (2.2) and (2.3) that

\[
\sum_{n=1}^{\infty} P \left( \sup_x |F_n(x) - F(x)| > D_0 \kappa_n \right) \leq C \sum_{n=1}^{\infty} \sum_{j=1}^{n-1} n^{-c_0 D_0^2/18} < \infty.
\]

This completes the proof of the lemma. \( \square \)
3. Main results and their proofs

In this section, we will provide the main results and their proofs. At first, we present the complete consistency and the convergence rate for the nearest neighbor density estimator \( f_n(x) \) of \( f(x) \).

THEOREM 3.1. Let \( \{X_n, n \geq 1\} \) be a sequence of WOD random variables. Suppose that \( k_n/n \to 0 \) and \( k_n^2/[n \log(nG)] \to \infty \) as \( n \to \infty \). Then for any \( x \in c(f) \),

\[
f_n(x) \to f(x) \text{ a.c., } n \to \infty.
\]

REMARK 3.1. If we take \( g(n) = O(n^\delta) \) in Theorem 3.1, where \( \delta \) is an arbitrary nonnegative constant, then Theorem 3.1 is equivalent to Corollary 2.1 of Wang and Hu \[10\]. Noticing that when \( g(n) \) grows geometrically, such as \( g(n) = O(e^{nt}) \) for some \( 0 < t < 1 \), the result of Theorem 3.1 still holds.

Proof. For any \( \varepsilon > 0 \), without loss of generality, assume that \( f(x) > \varepsilon \). For any \( x \in c(f) \), denote

\[
b_n(x) = \frac{k_n}{2n(f(x) + \varepsilon)}, \quad c_n(x) = \frac{k_n}{2n(f(x) - \varepsilon/2)}.
\]

It follows by (1.1) that

\[
A_x =: \{|f_n(x) - f(x)| > \varepsilon\} = \{f_n(x) > f(x) + \varepsilon\} \bigcup \{f_n(x) < f(x) - \varepsilon, f(x) > \varepsilon\} \subset \{f_n(x) > f(x) + \varepsilon\} \bigcup \{f_n(x) < f(x) - \varepsilon/2, f(x) > \varepsilon\} = \{a_n(x) < b_n(x)\} \bigcup \{a_n(x) > c_n(x), f(x) > \varepsilon\} \subset \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) \geq \frac{k_n}{n} \right\} \bigcup \left\{ F_n(x + c_n(x)) - F_n(x - c_n(x)) \leq \frac{k_n}{n}, f(x) > \varepsilon \right\} =: A_{1x} \bigcup A_{2x}.
\]

(3.1)

Noting that \( F'(x) = f(x) \), \( \lim_{n \to \infty} b_n = 0 \) and \( \lim_{n \to \infty} c_n = 0 \), we obtain

\[
\lim_{n \to \infty} \frac{F(x + b_n(x)) - F(x - b_n(x))}{2b_n(x)} = f(x), \quad \lim_{n \to \infty} \frac{F(x + c_n(x)) - F(x - c_n(x))}{2c_n(x)} = f(x),
\]

which imply that for all sufficiently large \( n \),

\[
F(x + b_n(x)) - F(x - b_n(x)) < 2b_n(x)(f(x) + \varepsilon/2) = \frac{k_n}{n} \frac{f(x) + \varepsilon/2}{f(x) + \varepsilon}
\]

(3.2)

and

\[
F(x + c_n(x)) - F(x - c_n(x)) > 2c_n(x)(f(x) + \varepsilon/4) = \frac{k_n}{n} \frac{f(x) - \varepsilon/4}{f(x) - \varepsilon/2}.
\]

(3.3)
Let $\delta(x) = \frac{\varepsilon}{8(f(x) + \varepsilon)} \leq \frac{1}{8}$. We have by (3.1) and (3.3) that
\[
A_{1x} = \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) \geq \frac{k_n}{n} \right\} 
\subset \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \geq \frac{k_n}{n} - \frac{f(x) + \varepsilon/2}{f(x) + \varepsilon} \right\}
\subset \left\{ |F_n(x + b_n(x)) - F(x + b_n(x))| + |F_n(x - b_n(x)) - F(x - b_n(x))| \geq 4\frac{k_n}{n}\delta(x) \right\}
\subset \left\{ \bigcup \left\{ |F_n(x - b_n(x)) - F(x - b_n(x))| \geq \frac{k_n}{n}\delta(x) \right\} \right\}
=: A_{11x} \bigcup A_{12x}.
\tag{3.4}
\]

Similarly, we have by (3.1) and (3.4) that
\[
A_{2x} = \left\{ F_n(x + c_n(x)) - F_n(x - c_n(x)) \leq \frac{k_n}{n}f(x) > \varepsilon \right\}
\subset \left\{ F_n(x + c_n(x)) - F_n(x - c_n(x)) - F(x + c_n(x)) + F(x - c_n(x)) \leq \frac{k_n}{n} - \frac{k_n}{n} \frac{f(x) - \varepsilon/4}{f(x) - \varepsilon/2}, f(x) > \varepsilon \right\}
\subset \left\{ |F_n(x + c_n(x)) - F(x + c_n(x))| + |F_n(x - c_n(x)) - F(x - c_n(x))| \geq \frac{k_n}{4n} \frac{\varepsilon}{f(x) - \varepsilon/2}, f(x) > \varepsilon \right\}
\subset \left\{ \bigcup \left\{ |F_n(x - c_n(x)) - F(x - c_n(x))| \geq \frac{k_n}{n}\delta(x) \right\} \right\}
=: A_{21x} \bigcup A_{22x}.
\tag{3.5}
\]

Hence, by (3.1), (3.4) and (3.5), we can get that
\[
A_x \subset A_{11x} \bigcup A_{12x} \bigcup A_{21x} \bigcup A_{22x}.
\tag{3.6}
\]

For fixed $x$, denote for $1 \leq i \leq n$ and $n \geq 1$ that
\[
X_{ni} = I(X_i < x + b_n(x)) - EI(X_i < x + b_n(x)).
\]
It follows by Lemma 2.1 again that $X_{n1}, X_{n2}, \ldots, X_{nn}$ are still WOD random variables, with $EX_{ni} = 0$ and $|X_{ni}| \leq 1$. Noting that $k_n/n \to 0$, $\delta(x) \leq \frac{1}{8}$, and applying Lemma 2.2 with $b_n = 1$ and $\Delta_n^2 = \sum_{i=1}^{n} EX_{ni}^2 \leq n$, we obtain that

$$P(A_{11x}) = P \left( \left| F(x + b_n(x)) - F(x + b_n(x)) \right| \geq \frac{k_n}{n} \delta(x) \right)$$

$$= P \left( \left| \sum_{i=1}^{n} X_{ni} \right| > k_n \delta(x) \right)$$

$$\leq Cg(n) \exp \left\{ -\frac{k_n^2 \delta^2(x)}{4 \Delta_n^2 + 2k_n \delta(x)} \right\}$$

$$\leq Cg(n) \exp \left\{ -\frac{k_n^2 \delta^2(x)}{5n} \right\}$$

$$\leq Cg(n) \exp \{-2 \log(ng(n))\} \leq Cn^{-2}. \quad (3.7)$$

Similarly, we can verify that (3.7) still holds for $A_{12x}$, $A_{21x}$ and $A_{22x}$. Hence, we have by (3.6) and (3.7) that

$$\sum_{n=1}^{\infty} P(|f_n(x) - f(x)| > \epsilon) = \sum_{n=1}^{\infty} P(A_x)$$

$$\leq \sum_{n=1}^{\infty} [P(A_{11x}) + P(A_{12x}) + P(A_{21x}) + P(A_{22x})]$$

$$\leq 4C \sum_{n=1}^{\infty} n^{-2} < \infty.$$

This completes the proof of the theorem. □

**Theorem 3.2.** Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables, and $f(x)$ satisfy the local Lipschitz condition at $x$ and $f(x) > 0$. If $k_n = O(n^{3/4} \log^{1/4}(ng(n)))$ and $\tau_n =: \sqrt{n \log(ng(n))}/k_n \to 0$ as $n \to \infty$, then for any sufficiently large $D > 0$,

$$\sum_{n=1}^{\infty} P(|f_n(x) - f(x)| > D \tau_n) < \infty,$$

and thus,

$$|f_n(x) - f(x)| \leq D \tau_n \ a.s., \ n \to \infty.$$

**Proof.** Noting that $f(x) > 0$ and $\tau_n \to 0$, we have $f(x) > D \tau_n$ for all sufficiently large $n$. Denote

$$\mu_n(x) = \frac{k_n}{2n(f(x) + D \tau_n)}, \text{ and } v_n(x) = \frac{k_n}{2n(f(x) - D \tau_n/2)}.$$
Similar to the proof of (3.1), we obtain

\[
B_x =: \{|f_n(x) - f(x)| > D\tau_n\} \\
= \{f_n(x) > f(x) + D\tau_n\} \cup \{f_n(x) < f(x) - D\tau_n, f(x) > D\tau_n\} \\
\subset \{f_n(x) > f(x) + D\tau_n\} \cup \{f_n(x) < f(x) - D\tau_n/2, f(x) > D\tau_n\} \\
= \{a_n(x) < \mu_n(x)\} \cup \{a_n(x) > \nu_n(x), f(x) > D\tau_n\} \\
\subset \left\{ F_n(x + \mu_n(x)) - F_n(x - \mu_n(x)) \geq \frac{k_n}{n} \right\} \\
\cup \left\{ F_n(x + \nu_n(x)) - F_n(x - \nu_n(x)) \leq \frac{k_n}{n}, f(x) > D\tau_n \right\} \\
=: B_{1x} \cup B_{2x}.
\]

(3.8)

By Differential Mean Value Theorem, we can see that there exist some \(\xi_{1n} \in (x - \mu_n(x), x + \mu_n(x))\) and \(\xi_{2n} \in (x - \nu_n(x), x + \nu_n(x))\) such that

\[
F(x + \mu_n(x)) - F(x - \mu_n(x)) = 2\mu_n(x)f(\xi_{1n}),
\]

and

\[
F(x + \nu_n(x)) - F(x - \nu_n(x)) = 2\nu_n(x)f(\xi_{2n}),
\]

and thus,

\[
F_n(x + \mu_n(x)) - F_n(x - \mu_n(x)) - F(x + \mu_n(x)) + F(x - \mu_n(x)) \\
\geq \frac{k_n}{n} - 2\mu_n(x)f(\xi_{1n}) = \frac{k_n}{n} \cdot \frac{f(x) - f(\xi_{1n}) + D\tau_n}{f(x) + D\tau_n}
\]

(3.11)

and

\[
F_n(x + \nu_n(x)) - F_n(x - \nu_n(x)) - F(x + \nu_n(x)) + F(x - \nu_n(x)) \\
\leq \frac{k_n}{n} - 2\nu_n(x)f(\xi_{2n}) = \frac{k_n}{n} \cdot \frac{f(x) - f(\xi_{2n}) - D\tau_n/2}{f(x) - D\tau_n/2}.
\]

(3.12)

Noting that \(k_n = O(n^{3/4} \log^{1/4}(ng(n)))\), we can see that there exists a positive constant \(c_1\) such that \(k_n \leq c_1 n^{3/4} \log^{1/4}(ng(n))\). Since \(f(x)\) satisfies the local Lipschitz condition at \(x\) and \(f(x) > 0\), \(\tau_n = \sqrt{n \log(ng(n))}/k_n \rightarrow 0\) as \(n \rightarrow \infty\), we can get that there exists a positive constant \(L(x)\) depending only on \(x\) such that for all sufficiently large \(n\),

\[
|f(x) - f(\xi_{1n})| \leq L(x)|x - \xi_{1n}| \leq L(x)\mu_n(x) \\
\leq \frac{L(x)k_n}{2nf(x)} = \frac{L(x)}{2f(x)} \cdot \tau_n \cdot \frac{k_n^2}{n^{3/2} \log^{1/2}(ng(n))} \leq \frac{c_1^2 L(x)}{2f(x)} \tau_n
\]

(3.13)
and
\[
|f(x) - f(\xi_{2n})| \leq L(x)|x - \xi_{2n}| \leq L(x)v_n(x) = \frac{L(x)k_n}{2n(f(x) - D\tau_n/2)} \leq \frac{L(x)k_n}{nf(x)} \leq \frac{c^2L(x)}{4f(x)} \tau_n.
\] (3.14)

Noting that \( f(x) \) is bounded, we define \( \sup_x f(x) = M < \infty \). Take \( D > \frac{c^2L(x)}{f(x)} \). It follows by (3.11)–(3.14) that
\[
\frac{k_n}{n} \cdot \frac{f(x) - f(\xi_{1n}) + D\tau_n}{f(x) + D\tau_n} \geq \frac{k_n}{n} \cdot \frac{c^2L(x)}{2f(x)} \tau_n + D\tau_n \geq \frac{k_n}{n} \cdot \frac{D}{4M}
\] (3.15)

and
\[
\frac{k_n}{n} \cdot \frac{f(x) - f(\xi_{2n}) - D\tau_n/2}{f(x) - D\tau_n/2} \leq \frac{k_n}{n} \cdot \frac{c^2L(x)}{2f(x)} \tau_n - D\tau_n/2 \leq -\frac{k_n}{n} \cdot \frac{D}{4M}.
\] (3.16)

Hence, we have by (3.11) and (3.15) that for all sufficiently large \( n \),
\[
B_{1x} \subset \left\{ F_n(x + \mu_n(x)) - F_n(x - \mu_n(x)) - F(x + \mu_n(x)) + F(x - \mu_n(x)) \geq \frac{k_n}{n} \cdot \frac{D}{4M} \right\}
\]
\[
\subset \left\{ |F_n(x + \mu_n(x)) - F(x + \mu_n(x))| \geq \frac{k_n}{n} \cdot \frac{D}{8M} \right\}
\]
\[
\cup \left\{ |F_n(x - \mu_n(x)) - F(x - \mu_n(x))| \geq \frac{k_n}{n} \cdot \frac{D}{8M} \right\}
\]
\[
=: B_{11x} \cup B_{12x}.
\] (3.17)

Similarly, we have by (3.12) and (3.16) that for all sufficiently large \( n \),
\[
B_{2x} \subset \left\{ |F_n(x + \nu_n(x)) - F(x + \nu_n(x))| \geq \frac{k_n}{n} \cdot \frac{D}{8M} \right\}
\]
\[
\cup \left\{ |F_n(x - \nu_n(x)) - F(x - \nu_n(x))| \geq \frac{k_n}{n} \cdot \frac{D}{8M} \right\}
\]
\[
=: B_{21x} \cup B_{22x}.
\] (3.18)

Therefore, we can get by (3.8), (3.17) and (3.18) that
\[
B_x \subset B_{11x} \cup B_{12x} \cup B_{21x} \cup B_{22x}.
\] (3.19)

For fixed \( x \), denote for \( 1 \leq i \leq n \) and \( n \geq 1 \) that
\[
X'_{ni} = I(X_i < x + \mu_n(x)) - EI(X_i < x + \mu_n(x)).
\]

It follows by Lemma 2.1 again that \( X'_{n1}, X'_{n2}, \ldots, X'_{nn} \) are still WOD random variables with \( EX'_{ni} = 0 \), and \( |X'_{ni}| \leq 1 \). Applying Lemma 2.2 with \( b_n = 1 \), \( \Delta^2 = \sum_{i=1}^n E(X'_{ni})^2 \leq \)
Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables, and $f(x)$ satisfy the local Lipschitz condition at $x$ and $f(x) > 0$. If $k_n = [n^{3/4} \log^{1/4}(ng(n))]$, then for all sufficiently large $D > 0$,

$$
\sum_{n=1}^{\infty} P\left(|f_n(x) - f(x)| > Dn^{-1/4} \log^{1/4}(ng(n))\right) < \infty,
$$

and thus,

$$
|f_n(x) - f(x)| \leq Dn^{-1/4} \log^{1/4}(ng(n)) \text{ a.s., } n \to \infty.
$$

**Remark 3.2.** Yang [9] and Wang and Hu [10] established the convergence rate of $o(n^{-1/4} \log^{1/4}(n \log \log n))$ a.s. for the nearest neighbor estimator based on NA samples and WOD samples, respectively. If we take $g(n) = O(n^\delta)$, where $\delta$ is an arbitrary nonnegative constant, then the convergence rate is of $O(n^{-1/4} \log^{1/4}(n))$ a.s. in Corollary 3.1, which still improves the corresponding ones of Yang [9] and Wang and Hu.
In addition, if \( g(n) \) grows geometrically, then the result of Wang and Hu [10] is invalid; However, we can still obtain a certain rate of convergence by Corollary 3.1 in this case. Hence, the result of Corollary 3.1 extends and improves the corresponding ones of Yang [9] and Wang and Hu [10].

In the following, we will present the uniformly complete consistency and its convergence rate for the nearest neighbor estimator.

**Theorem 3.3.** Let \( \{X_n, n \geq 1\} \) be a sequence of WOD random variables, and \( f(x) \) be uniformly continuous. If \( k_n/n \to 0 \) and \( k_n^2/[n \log(n g(n))] \to \infty \) as \( n \to \infty \), then for any \( \varepsilon > 0 \),

\[
\sum_{n=1}^{\infty} P \left( \sup_{x} |f_n(x) - f(x)| > \varepsilon \right) < \infty,
\]

and thus,

\[
\sup_{x} |f_n(x) - f(x)| \to 0 \text{ a.s., } n \to \infty.
\]

**Proof.** We use the same notations as those in Theorem 3.1. Since the density function \( f(x) \) is uniformly continuous, for any \( \varepsilon > 0 \), there exist a positive constant \( \delta_0 \) such that

\[
|f(x) - f(y)| < \frac{\varepsilon}{4},
\]

when \( |x - y| < \delta_0 \). Noting that \( k_n/n \to 0 \) as \( n \to \infty \), we can get that for any \( x \) and all sufficiently large \( n \),

\[
b_n(x) = \frac{k_n}{2n(f(x) + \varepsilon)} < \delta_0, \quad \text{and} \quad c_n(x) = \frac{k_n}{2n(f(x) - \varepsilon/2)} < \delta_0.
\]

By Differential Mean Value Theorem, we can see that there exist \( \eta_{1n} \in (x - b_n(x), x + b_n(x)) \) and \( \eta_{2n} \in (x - c_n(x), x + c_n(x)) \) such that

\[
F(x + b_n(x)) - F(x - b_n(x)) = 2b_n(x)f(\eta_{1n}),
\]

and

\[
F(x + c_n(x)) - F(x - c_n(x)) = 2c_n(x)f(\eta_{2n}).
\]

Hence, it follows by (3.22) that \( |x - \eta_{1n}| < \delta_0 \) and \( |x - \eta_{2n}| < \delta_0 \), which together with (3.21) yields that

\[
|f(x) - f(\eta_{1n})| < \frac{\varepsilon}{4}, \quad \text{and} \quad |f(x) - f(\eta_{2n})| < \frac{\varepsilon}{4}.
\]
Denote \( \sup_x f(x) = M < \infty \), \( \delta(M) = \frac{\epsilon}{8(M + \epsilon)} \) and \( A = \{ \sup_x |F_n(x) - F(x)| \geq \delta(M) \frac{kn}{n} \} \). By (3.23) and (3.25), we obtain

\[
A_{1x} = \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \right\} \\
\quad \geq \frac{k_n}{n} - 2b_n(x)f(\eta_{1n}) \}
\]

\[
= \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \right\} \\
\quad \geq \frac{k_n}{n} \cdot \frac{f(x) - f(\eta_{1n}) + \epsilon}{f(x) + \epsilon}
\]

\[
\subset \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \geq \frac{k_n}{n} \cdot \frac{-\epsilon/4 + \epsilon}{f(x) + \epsilon} \right\}
\]

\[
\subset \left\{ F_n(x + b_n(x)) - F_n(x - b_n(x)) - F(x + b_n(x)) + F(x - b_n(x)) \geq \frac{k_n}{n} \cdot 2\delta(M) \right\}
\]

\[
\subset \left\{ |F_n(x + b_n(x)) - F(x + b_n(x))| \geq \frac{k_n}{n} \delta(M) \right\}
\]

\[
\bigcup \left\{ |F_n(x - b_n(x)) - F(x - b_n(x))| \geq \frac{k_n}{n} \delta(M) \right\} \subset A. \tag{3.26}
\]

Similarly, we have by (3.24) and (3.25) that

\[
A_{2x} = \left\{ F_n(x + c_n(x)) - F_n(x - c_n(x)) - F(x + c_n(x)) + F(x - c_n(x)) \right\} \\
\quad \leq \frac{k_n}{n} \cdot \frac{f(x) - f(\eta_{2n}) - \epsilon/2}{f(x) - \epsilon/2}, f(x) > \epsilon \}
\]

\[
\subset \left\{ F_n(x + c_n(x)) - F_n(x - c_n(x)) - F(x + c_n(x)) + F(x - c_n(x)) \right\} \\
\quad \leq \frac{k_n}{n} \cdot \frac{\epsilon/4 - \epsilon/2}{f(x) + \epsilon}
\]

\[
\subset \left\{ |F_n(x + c_n(x)) - F(x + c_n(x))| + |F_n(x - c_n(x)) - F(x - c_n(x))| \right\} \\
\quad \leq -\frac{k_n}{n} \cdot 2\delta(M)
\]

\[
\subset \left\{ |F_n(x + c_n(x)) - F(x + c_n(x))| \geq \frac{k_n}{n} \delta(M) \right\}
\]

\[
\bigcup \left\{ |F_n(x - c_n(x)) - F(x - c_n(x))| \geq \frac{k_n}{n} \delta(M) \right\} \subset A. \tag{3.27}
\]

Noting that \( k_n^2/(n \log(n)) \rightarrow \infty \) as \( n \rightarrow \infty \), we can obtained that for all sufficiently large \( n \), \( \frac{k_n}{n} \delta(M) \geq D_0(\log(n))/n^{1/2} \). On the other hand, it follows from (3.1), (3.26) and (3.27) that \( A_x \subset A \) for any \( x \). Hence, applying Lemma 2.4 with
DENSITY FUNCTION BASED ON WOD SAMPLES

\[ \kappa_n = \left( \log(ng(n)) / n \right)^{1/2}, \]
we can get

\[ \sum_{n=1}^{\infty} P \left( \sup_x |f_n(x) - f(x)| > \varepsilon \right) = \sum_{n=1}^{\infty} P \left( \bigcup_x A_x \right) \]
\[ \leq \sum_{n=1}^{\infty} P \left( \sup_x |F_n(x) - F(x)| \geq \frac{k_n}{n} \delta(M) \right) \]
\[ \leq \sum_{n=1}^{\infty} P \left( \sup_x |F_n(x) - F(x)| \geq D_0 \left( \log(ng(n)) / n \right)^{1/2} \right) \]
\[ < \infty. \]

This completes the proof of the theorem. \[\square\]

**THEOREM 3.4.** Let \( \{X_n, n \geq 1\} \) be a sequence of WOD random variables, and \( f(x) \) satisfy the Lipschitz condition on \( \mathbb{R} \). If \( k_n = O(n^{2/3} \log^{1/3}(ng(n))) \) and \( \tau_n =: \sqrt{n \log(ng(n)) / k_n} \to 0 \) as \( n \to \infty \), then for all sufficiently large \( D > 0 \),

\[ \sum_{n=1}^{\infty} P \left( \sup_x |f_n(x) - f(x)| > D \tau_n \right) < \infty, \]

and thus,

\[ \sup_x |f_n(x) - f(x)| \leq D \tau_n \text{ a.s., } n \to \infty. \]

**Proof:** We use the same notations as those in Theorem 3.2. Noting that \( k_n = O(n^{2/3} \log^{1/3}(ng(n))) \), there exists a positive constant \( c_2 \) such that

\[ k_n \leq c_2 n^{2/3} \log^{1/3}(ng(n)). \]

Since the density function \( f(x) \) satisfies the Lipschitz condition on \( \mathbb{R} \), and

\[ \tau_n = \sqrt{n \log(ng(n)) / k_n} \to 0 \quad \text{as} \quad n \to \infty, \]

there exists a positive constant \( L \) depending not on \( x \) such that for all sufficiently large \( n \),

\[ |f(x) - f(\xi_{1n})| \leq L |x - \xi_{1n}| \leq L \mu_n(x) \leq \frac{Lk_n}{2D \tau_n} \]
\[ = \frac{L}{2D} \cdot \frac{\sqrt{n \log(ng(n))}}{k_n} \cdot \frac{k_n^3}{n^2 \log(ng(n))} \leq \frac{c_2 L}{2D} \tau_n \quad (3.28) \]

and

\[ |f(x) - f(\xi_{2n})| \leq L |x - \xi_{2n}| \leq L \nu_n(x) \]
\[ \leq \frac{Lk_n}{2n(D \tau_n - D \tau_n / 2)} \leq \frac{Lk_n}{D n \tau_n} \leq \frac{c_2 L}{D} \tau_n. \quad (3.29) \]
Denote $\sup_x f(x) = M < \infty$. Taking sufficiently large $D$ such that $D > \frac{4c^3 L}{D}$, we have by (3.11), (3.12), (3.28) and (3.29) that

$$\frac{k_n}{n} \cdot \frac{f(x) - f(\xi_{1n}) + D \tau_n}{f(x) + D \tau_n} \geq \frac{k_n}{n} \cdot \frac{-cL}{2D} \tau_n + D \tau_n \geq \frac{k_n}{n} \cdot D \tau_n \cdot \frac{D}{4M}$$

(3.30)

and

$$\frac{k_n}{n} \cdot \frac{f(x) - f(\xi_{2n}) - D \tau_n/2}{f(x) - D \tau_n/2} \leq \frac{k_n}{n} \cdot \frac{cL}{2D} \tau_n - D \tau_n/2 \leq -\frac{k_n}{n} \cdot \frac{D}{4M}.$$  

(3.31)

Denote $B = \{ \sup_x |F_n(x) - F(x)| \geq \frac{k_n \tau_n}{n} \cdot \frac{D}{8M} \}$. It is easily checked that for all sufficiently large $n$,

$$B_{1x} \subset \left\{ |F_n(x + \mu_n(x)) - F(x + \mu_n(x))| \geq \frac{k_n \tau_n}{n} \cdot \frac{D}{8M} \right\}$$

$$\bigcup \left\{ |F_n(x - \mu_n(x)) - F(x - \mu_n(x))| \geq \frac{k_n \tau_n}{n} \cdot \frac{D}{8M} \right\} \subset B.$$  

(3.32)

Similarly, we can also obtain that for all sufficiently large $n$,

$$B_{2x} \subset \left\{ |F_n(x + \nu_n(x)) - F(x + \nu_n(x))| \geq \frac{k_n \tau_n}{n} \cdot \frac{D}{8M} \right\}$$

$$\bigcup \left\{ |F_n(x - \nu_n(x)) - F(x - \nu_n(x))| \geq \frac{k_n \tau_n}{n} \cdot \frac{D}{8M} \right\} \subset B.$$  

(3.33)

Hence, it follows from (3.8), (3.32) and (3.33) that $B_x \subset B$ for any $x$. Taking $\kappa_n = k_n \tau_n = (\log(n/\gamma(n)))^{1/2}$ in Lemma 2.4, and for all sufficiently large $D$ such that $\frac{D}{8M} \geq D_0$, we get

$$\sum_{n=1}^\infty P\left( \sup_x |f_n(x) - f(x)| > D \tau_n \right) = \sum_{n=1}^\infty P\left( \bigcup_{x} B_x \right)$$

$$\leq \sum_{n=1}^\infty P\left( \sup_x |F_n(x) - F(x)| \geq \frac{k_n \tau_n}{n} \cdot \frac{D}{8M} \right)$$

$$\leq \sum_{n=1}^\infty P\left( \sup_x |F_n(x) - F(x)| \geq D_0(\log n/n)^{1/2} \right)$$

$$< \infty,$$

which implies the desired result immediately. The proof is completed. \hfill \Box

Taking $k_n = \lfloor n^{2/3} \log^{1/3}(n/\gamma(n)) \rfloor$ in Theorem 3.4, we can get the following result.

**Corollary 3.2.** Let $\{X_n, n \geq 1\}$ be a sequence of WOD random variables, and density function $f(x)$ satisfy the Lipschitz condition on $\mathbb{R}$. If $k_n = \lfloor n^{2/3} \log^{1/3}(n/\gamma(n)) \rfloor$, then for any sufficiently large $D > 0$,

$$\sum_{n=1}^\infty P\left( \sup_x |f_n(x) - f(x)| > Dn^{-1/6} \log^{1/6}(n/\gamma(n)) \right) < \infty,$$
and thus,
\[
\sup_x |f_n(x) - f(x)| \leq D n^{-1/6} \log^{1/6}(ng(n)) \quad \text{a.s., } n \to \infty.
\]

**Remark 3.3.** Yang [9] and Wang and Hu [10] established the uniformly strong convergence rate of \( o(n^{-1/6} \log^{1/6} n \log \log n) \) a.s. for the nearest neighbor estimator based on NA and WOD samples, respectively. When \( g(n) \) grows polynomially, the convergence rate obtained in Corollary 3.2 is slightly faster than their results; when \( g(n) \) grows geometrically, our result is still valid. Hence, our result extends and improves the corresponding ones of Yang [9] and Wang and Hu [10].

**References**


(Received July 22, 2020)

Kan Chen
School of Mathematics and Statistics
Chaohu University
Hefei, 238024, P.R. China
e-mail: kanchenchu@126.com

Xufei Tang
School of Mathematics and Statistics
Chaohu University
Hefei, 238024, P.R. China

Jiangfeng Hao
School of Mathematics and Statistics
Chaohu University
Hefei, 238024, P.R. China

Journal of Mathematical Inequalities
www.ele-math.com
jmi@ele-math.com