PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 40, Fasc. 1 (2020), pp. 83–95
Published online 20.3.2020
doi:10.37190/0208-4147.40.1.5

MODERATE DEVIATION AND LARGE DEVIATION FOR WEGMAN–DAVIES RECURSIVE DENSITY ESTIMATORS

BY

YU MIAO* (XINXIANG), QINGHUI GAO (XINXIANG),
JIANYONG MU (XINXIANG), AND CONGHUI DENG (XINXIANG)

Abstract. Let $\{X_k, k \geqslant 1\}$ be a sequence of independent identically distributed random variables with common probability density function f, and let $\hat{f}_n$ denote a Wegman–Davies recursive density estimator

$$\hat{f}_n(x) = \frac{1}{nh_n^{1/2}} \sum_{j=1}^{n} \frac{1}{h_j^{1/2}} K\left(\frac{x - X_j}{h_j}\right)$$

where K is a kernel function and h_n is a band sequence. In the present paper, the moderate deviation principle and the large deviation principle for the estimator $\hat{f}_n$ are established.

2020 Mathematics Subject Classification: Primary 62G07; Secondary 60F10.

Key words and phrases: moderate deviation principle, large deviation principle, recursive kernel estimator.

1. INTRODUCTION

Given a sequence of independent identically distributed random variables $X_1, X_2, \ldots$ with common probability density function $f(x)$, how can one estimate $f(x)$? This problem has been of long-lasting interest among statisticians, and numerous interesting and fundamental results have been obtained.

1.1. Rosenblatt estimator. Rosenblatt [15] introduced the following kernel estimator of the density $f(x)$:

$$f_n^*(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right),$$

and Parzen [14] studied many important properties of these estimators, such as

* This work is supported by NSFC (11471104, 11971154).

consistency, asymptotic normality, uniform consistency, etc. Csörgö and Horváth [1] considered the central limit theorem for the L_p-norm of f_n^*. Lu [9] studied the kernel methods for density estimation of stationary samples under generalized conditions, which unify both the linear and α-mixing processes and also apply to the non-linear and/or non-α-mixing processes. Furthermore, under general, mild conditions, the estimators f_n^* were shown to be asymptotically normal. Louani [8] established a large deviation limit theorem of Chernoff type for f_n^* in the L_1-distance. Louani [7] and Gao [3] obtained the large deviation and moderate deviation principles for f_n^* for pointwise and L_∞ convergence. Mokkadem et al. [13] studied the large and moderate deviation principles for kernel estimators of the partial derivatives of f. Gao [4] proved the moderate deviation principle and the law of the iterated logarithm in $L_1(\mathbb{R}^d)$ for f_n^*.

1.2. Wolverton–Wagner estimator or Yamato estimator. It is well known that the Rosenblatt estimator f_n^* is a non-recursive kernel density estimator. Wolverton and Wagner [17] defined a related estimator

$$\tilde{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{h_j}K\left(\frac{x-X_j}{h_j}\right),$$

which was apparently independently introduced by Yamato [18]. The estimator $\tilde{f}_n(x)$ is useful in practice, because it can be calculated recursively,

$$\tilde{f}_n(x) = \frac{n-1}{n}\tilde{f}_{n-1}(x) + \frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right).$$

Masry and Györfi [11] obtained the sharp rates for almost sure convergence of $\tilde{f}_n$ when the process $\{X_i, i \geqslant 1\}$ is asymptotically uncorrelated. Liang and Baek [5], [6] discussed the point asymptotic normality and the Berry–Esseen bounds of $\tilde{f}_n$ for strictly stationary samples of negatively associated random variables. Mokkadem et al. [12] obtained the large and moderate deviation principles for the recursive kernel estimator of the probability density function $\tilde{f}_n$ and its partial derivatives in the multivariate case.

1.3. Wegman–Davies estimator. Wegman and Davies [16] introduced another related estimator $\hat{f}_n(x)$ by

$$\hat{f}_n(x) = \frac{1}{nh_n^{1/2}}\sum_{j=1}^{n}\frac{1}{h_j^{1/2}}K\left(\frac{x-X_j}{h_j}\right),$$

which can also be calculated recursively:

$$\hat{f}_n(x) = \frac{n-1}{n}\left(\frac{h_{n-1}}{h_n}\right)^{1/2}\hat{f}_{n-1}(x) + \frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right).$$

Masry and Györfi [11] established the sharp rates for almost sure convergence of $\hat{f}_n$ when the process $\{X_i, i \geqslant 1\}$ is asymptotically uncorrelated. Liang and Baek [6] discussed the Berry–Esseen bounds of $\hat{f}_n$ for strictly stationary samples of negatively associated random variables. Zhang and Liang [19] obtained the point asymptotic normality of $\hat{f}_n$ for such samples.

Motivated by that work, in the present paper we shall discuss the large deviation and the moderate deviation behaviour of the estimators $\hat{f}_n$.

2. MODERATE DEVIATION PRINCIPLE

We assume that K is a Borel function satisfying

$$\|K\|_\infty := \sup_{y\in\mathbb{R}} |K(y)| < \infty, \quad \int_{\mathbb{R}} |K(y)|\, dy < \infty, \quad \lim_{y\to\pm\infty} |yK(y)| = 0, \tag{2.1}$$

and that $\{h_n\}$ is a sequence of positive real numbers satisfying

$$h_n \downarrow 0, \quad nh_n \to \infty. \tag{2.2}$$

Other assumptions about K and $\{h_n\}$ will be made as needed. Firstly, we recall the following theorem of Wegman and Davies [16].

THEOREM 2.1 ([16, Theorem 1]). *Let K and $\{h_n\}$ satisfy* (2.1) *and* (2.2).

(a) *If f is continuous at x, then*

$$nh_n \operatorname{Var}(\hat{f}_n(x)) \to f(x) \int_{\mathbb{R}} K^2(u)\, du.$$

(b) *Let*

$$K^*(u) = \int_{\mathbb{R}} e^{-iuy} K(y)\, dy$$

be the Fourier transform of K. Suppose that for some positive integer β,

$$\lim_{u\to 0} [1 - K^*(u)]/|u|^\beta = k_\beta$$

is finite and the derivative $f^{(\beta)}(x)$ of order β at x exists. Suppose finally that

$$nh_n^{\beta+1/2} \to \infty \quad \text{and} \quad \frac{1}{nh_n^{\beta+1/2}} \sum_{j=1}^n h_j^{\beta+1/2} \to \gamma_{\beta+1/2}. \tag{2.3}$$

Then

$$\frac{\mathbb{E}\hat{f}_n(x) - f(x) n^{-1} h_n^{-1/2} \sum_{j=1}^n h_j^{1/2}}{h_n^\beta} \to \gamma_{\beta+1/2} \cdot k_\beta \cdot f^{(\beta)}(x).$$

REMARK 2.1. If there exists a value of β such that k_β is non-zero, it is called the *characteristic exponent* of $K^*(u)$, and k_β is the *characteristic coefficient*.

THEOREM 2.2. *Let K and $\{h_n\}$ satisfy* (2.1) *and* (2.2). *Assume that a sequence $\{b_n\}$ of positive real numbers satisfies*

$$\lim_{n\to\infty} b_n = \infty, \qquad \lim_{n\to\infty} \frac{b_n}{\sqrt{nh_n}} = 0. \tag{2.4}$$

If f is continuous at x, then for any $r > 0$,

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{nh_n}}{b_n} |\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \geqslant r \right) = -\frac{r^2}{2\sigma^2(x)}$$

where

$$\sigma^2(x) = f(x) \int_{-\infty}^{\infty} K^2(u)\, du.$$

Proof. For any $r > 0$, we will prove

$$\frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{nh_n}}{b_n} (\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)) \geqslant r \right) \to -\frac{r^2}{2\sigma^2(x)};$$

the proof of

$$\frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{nh_n}}{b_n} (\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)) \leqslant -r \right) \to -\frac{r^2}{2\sigma^2(x)}$$

is similar. For any $\lambda \in \mathbb{R}$, let

$$\Lambda(\lambda) := \frac{\lambda^2\sigma^2(x)}{2}.$$

The Fenchel–Legendre transform of $\Lambda(\cdot)$ is

$$\Lambda^*(r) := \sup_{\lambda\in\mathbb{R}}\{\lambda r - \Lambda(\lambda)\} = \sup_{\lambda\in\mathbb{R}}\left\{ \lambda r - \frac{\lambda^2\sigma^2(x)}{2} \right\} = \frac{r^2}{2\sigma^2(x)}, \qquad r \in \mathbb{R}.$$

Let $Z_n := \frac{\sqrt{nh_n}}{b_n}(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x))$, and let $\Lambda_n(\cdot)$ be the logarithmic moment generating function of Z_n. By the Gärtner–Ellis theorem [2, Theorem 2.3.6], it is enough to show that for any $\lambda \in \mathbb{R}$,

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{b_n^2} \Lambda_n(\lambda b_n^2 Z_n). \tag{2.5}$$

Define

$$Y_j := h_j^{-1/2}\left(K\left(\frac{x - X_j}{h_j}\right) - \mathbb{E}K\left(\frac{x - X_j}{h_j}\right) \right).$$

Then we have

$$\begin{aligned}
\frac{1}{b_n^2}\Lambda_n(\lambda b_n^2 Z_n) &= \frac{1}{b_n^2}\log\mathbb{E}\exp\big(\lambda b_n\sqrt{nh_n}(\hat f_n(x)-\mathbb{E}\hat f_n(x))\big)\\
&= \frac{1}{b_n^2}\sum_{j=1}^n \log\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}h_j^{-1/2}\bigg(K\bigg(\frac{x-X_j}{h_j}\bigg)-\mathbb{E}K\bigg(\frac{x-X_j}{h_j}\bigg)\bigg)\bigg)\\
&= \frac{1}{b_n^2}\sum_{j=1}^n \log\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)
\end{aligned}$$

and

$$\begin{aligned}
&\bigg|\frac{1}{b_n^2}\sum_{j=1}^n \log\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-\frac{\lambda^2\sigma^2(x)}{2}\bigg|\\
&\qquad\leqslant \frac{1}{b_n^2}\sum_{j=1}^n\bigg|\log\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-\bigg[\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-1\bigg]\bigg|\\
&\qquad\quad+\frac{1}{b_n^2}\bigg|\sum_{j=1}^n\bigg[\bigg(\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-1\bigg)-\frac{\lambda^2 b_n^2\sigma^2(x)}{2n}\bigg]\bigg|\\
&\qquad=: \triangle_1+\triangle_2.
\end{aligned}$$

Now we need to control the terms $\triangle_1$ and $\triangle_2$. For $\triangle_2$ we can write

$$\begin{aligned}
\triangle_2 &= \frac{1}{b_n^2}\bigg|\sum_{j=1}^n\bigg[\bigg(\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-1\bigg)-\frac{\lambda^2 b_n^2\sigma^2(x)}{2n}\bigg]\bigg|\\
&\leqslant \frac{1}{b_n^2}\sum_{j=1}^n\bigg|\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-1-\frac{\lambda^2 b_n^2}{2n}\mathbb{E}Y_j^2\bigg|\\
&\quad+\frac{1}{b_n^2}\bigg|\sum_{j=1}^n\frac{\lambda^2 b_n^2}{2n}\mathbb{E}Y_j^2-\frac{\lambda^2 b_n^2\sigma^2(x)}{2n}\bigg|\\
&=: \triangle_{21}+\triangle_{22}.
\end{aligned}$$

For the term $\triangle_{21}$, by using the elementary inequality

$$\bigg|e^x-1-x-\frac{x^2}{2}\bigg|\leqslant\frac{|x|^3}{6}e^{|x|}\quad\text{for all } x\in\mathbb{R},$$

the condition (2.4) and the fact that $K(\cdot)$ is a bounded Borel function (which implies $|Y_i|\leqslant 2h_i^{-1/2}\|K\|_\infty$), for any n large enough we get

$$\begin{aligned}
\triangle_{21} &= \frac{1}{b_n^2}\sum_{j=1}^n\bigg|\mathbb{E}\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}Y_j\bigg)-1-\frac{\lambda^2 b_n^2}{2n}\mathbb{E}Y_j^2\bigg|\\
&\leqslant \frac{1}{b_n^2}\sum_{j=1}^n\frac{1}{6}\frac{\lambda^3 b_n^3}{n^{3/2}}\mathbb{E}|Y_j^3|\exp\bigg(\frac{\lambda b_n}{\sqrt{n}}|Y_j|\bigg)\\
&\leqslant C_0\frac{b_n\lambda^3}{n^{1/2}}\frac{1}{n}\sum_{j=1}^n\mathbb{E}|Y_j|^3\leqslant C_1\frac{b_n\lambda^3}{n^{1/2}h_n^{1/2}}\frac{1}{n}\sum_{j=1}^n\mathbb{E}Y_j^2
\end{aligned}$$

where C_0, C_1 are positive constants depending on the function $K(\cdot)$. So by using (2.4) again, and the fact (Theorem 2.1) that

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}Y_j^2 = \frac{1}{n}\sum_{j=1}^{n} h_j^{-1}\mathbb{E}\Big(K\Big(\frac{x-X_j}{h_j}\Big) - \mathbb{E}K\Big(\frac{x-X_j}{h_j}\Big)\Big)^2 = nh_n \operatorname{Var}(\hat{f}_n(x)) \to \sigma^2(x), \tag{2.6}$$

we have $\triangle_{21} \to 0$. Next we consider the term $\triangle_{22}$. From (2.6), we have

$$\triangle_{22} = \frac{\lambda^2}{2}\left|\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}Y_j^2 - \sigma^2(x)\right| \to 0.$$

For the term $\triangle_1$, by the elementary inequality

$$|e^x - 1 - x| \leqslant \frac{x^2}{2}e^{|x|} \quad \text{for all } x \in \mathbb{R},$$

and (2.4), for all n large enough we have

$$\sup_{1\leqslant j\leqslant n}\left|\mathbb{E}\exp\Big(\lambda\frac{b_n}{\sqrt{n}}Y_j\Big) - 1\right| \leqslant C_2\frac{\lambda^2 b_n^2}{2n}\sup_{1\leqslant j\leqslant n}\mathbb{E}Y_j^2 \leqslant C_3\frac{\lambda^2 b_n^2}{nh_n} \to 0 \tag{2.7}$$

where C_2 and C_3 are positive constants depending on $K(\cdot)$. Therefore for all n large enough,

$$\sup_{1\leqslant j\leqslant n}\left|\mathbb{E}\exp\Big(\lambda\frac{b_n}{\sqrt{n}}Y_j\Big) - 1\right| \leqslant \frac{1}{2}. \tag{2.8}$$

Using the elementary inequality

$$|\log(1+x) - x| \leqslant 2x^2 \quad \text{for } |x| \leqslant 1/2,$$

and the conditions (2.7) and (2.8), for any n large enough we have

$$\begin{aligned}
\triangle_1 &= \frac{1}{b_n^2}\sum_{j=1}^{n}\left|\log\mathbb{E}\exp\Big(\frac{\lambda b_n}{\sqrt{n}}Y_j\Big) - \Big[\mathbb{E}\exp\Big(\frac{\lambda b_n}{\sqrt{n}}Y_j\Big) - 1\Big]\right| \\
&\leqslant \frac{2}{b_n^2}\sum_{j=1}^{n}\Big(\mathbb{E}\exp\Big(\frac{\lambda b_n}{\sqrt{n}}Y_j\Big) - 1\Big)^2 \\
&\leqslant C_4\frac{1}{b_n^2}\sum_{j=1}^{n}\frac{\lambda^4 b_n^4}{n^2}(\mathbb{E}Y_j^2)^2 \leqslant C_5\frac{\lambda^4 b_n^2}{nh_n}\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}Y_j^2
\end{aligned}$$

where C_4 and C_5 are positive constants depending on $K(\cdot)$. By (2.4) and (2.6), we have $\triangle_1 \to 0$. From the above discussion, we obtain the desired result (2.5). ∎

COROLLARY 2.1. *Under the assumptions in Theorems* 2.1 *and* 2.2, *assume further that*

$$\frac{\sqrt{n}\, h_n^{1/2+\beta}}{b_n} \to 0. \tag{2.9}$$

Then for any $r > 0$, *we have*

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{nh_n}}{b_n} |\hat{f}_n(x) - \gamma_n f(x)| \geqslant r \right) = -\frac{r^2}{2\sigma^2(x)}$$

where $\gamma_n = n^{-1} h_n^{-1/2} \sum_{j=1}^n h_j^{1/2}$.

Proof. From Theorems 2.1 and 2.2, it is enough to show

$$\frac{\sqrt{nh_n}}{b_n} |\mathbb{E}\hat{f}_n(x) - \gamma_n f(x)| \to 0;$$

but the condition (2.9) guarantees this. ∎ The recursive estimator $\hat{f}_n(x)$ is not an (asymptotically) unbiased estimator of $f(x)$, but $\gamma_n^{-1}\hat{f}_n(x)$ is asymptotically unbiased. Hence we have the following corollary.

COROLLARY 2.2. *Under the assumptions in Theorems* 2.1 *and* 2.2, *set* $h_n = n^{-\gamma}$, *where* $0 < \gamma < 1$. *Furthermore, for any* $\beta > 0$ *such that* $\gamma(\beta + 1/2) < 1/2$, *take* b_n *satisfying* $b_n^{-1} = o(n^{\gamma(\beta+1/2)-1/2})$ *and* $b_n = o(n^{(1-\gamma)/2})$. *Then for any* $r > 0$, *we have*

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n^{1-\gamma}}}{b_n} |\gamma_n^{-1}\hat{f}_n(x) - f(x)| \geqslant r \right) = -\frac{r^2(1-\gamma/2)^2}{2\sigma^2(x)}$$

where $\gamma_n = n^{-1} h_n^{-1/2} \sum_{j=1}^n h_j^{1/2}$.

Proof. It is easy to see that the conditions (2.3), (2.4) and (2.9) hold and

$$\gamma_n \to \frac{1}{1-\gamma/2}.$$

So from Corollary 2.1, we get the desired result. ∎

REMARK 2.2. Assume that for some $l > 0$,

$$\frac{1}{n} \sum_{j=1}^n (h_j/h_n)^l \to \beta_l \quad \text{as } n \to \infty.$$

Masry [10] pointed out that under some conditions the recursive estimator $\hat{f}_n(x)$ is not an asymptotically unbiased estimator of $f(x)$. However, it is clear that after a simple scaling,

$$\hat{\hat{f}}_n(x) = \frac{\hat{f}_n(x)}{\beta_{1/2}},$$

$\hat{\hat{f}}_n(x)$ is an asymptotically unbiased estimator of $f(x)$. When $K(\cdot)$ is even symmetric, the dominant terms of the bias of $\tilde{f}_n(x)$ and $\hat{\hat{f}}_n(x)$ are

$$\mathrm{bias}[\hat{\hat{f}}_n(x)] \sim \frac{c_2\beta_{2.5}f^{(2)}(x)}{2\beta_{0.5}}h_n^2,$$
$$\mathrm{bias}[\tilde{f}_n(x)] \sim \frac{c_2\beta_2 f^{(2)}(x)}{2}h_n^2.$$

If we take $h_n = n^{-\nu}$, then $\tilde{f}_n(x)$ will generally have smaller bias than that of $\hat{\hat{f}}_n(x)$. Furthermore, Masry [10] showed that $\tilde{f}_n(x)$ has a larger variance than $\hat{\hat{f}}_n(x)$ under the strong mixing condition for the samples $\{X_i, i \geqslant 1\}$.

3. LARGE DEVIATION PRINCIPLE

In this section, we assume that K is a Borel function and the following conditions are satisfied:

(L1) The density function f is bounded and $K(\cdot)$ is a non-negative function such that for any $t \geqslant 0$,

$$\int_{\mathbb{R}} (e^{tK(z)} - 1)\, dz < \infty. \tag{3.1}$$

(L1′) $K(\cdot)$ is a non-negative function such that for any $t \geqslant 0$,

$$\int_{\mathbb{R}} (e^{tK(z)} - 1)\, dz < \infty, \qquad \lim_{|z|\to\infty} |z|(e^{tK(z)} - 1) = 0.$$

(L2) $h_n = n^{-\alpha}$ with $0 < \alpha < 1$.

To state the large deviation principle for the estimator $\hat{f}_n$, we need the following lemmas. Their proofs are elementary, but we give them for completeness.

LEMMA 3.1. *Let a_n be a sequence of positive numbers such that $a_n \to 0$. If* **(L1)** *or* **(L1′)** *holds, then for each continuity point x of $f(\cdot)$ and for any $t > 0$, we have*

$$\lim_{n\to\infty} \int_{\mathbb{R}} (e^{tK(z)} - 1) f(x - a_n z)\, dz = f(x) \int_{\mathbb{R}} (e^{tK(z)} - 1)\, dz, \tag{3.2}$$

$$\lim_{n\to\infty} \int_{\mathbb{R}} K(z) f(x - a_n z)\, dz = f(x) \int_{\mathbb{R}} K(z)\, dz. \tag{3.3}$$

Proof. (1) Assume that **(L1)** is satisfied. By the dominated convergence theorem, the claim (3.2) holds. Furthermore, by using the inequality $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$, we have

$$\int_{\mathbb{R}} K(z)\, dz \leqslant \int_{\mathbb{R}} (e^{K(z)} - 1)\, dz < \infty,$$

which implies (3.3) by the dominated convergence theorem.

(2) Assume that the condition **(L1′)** is satisfied. For any $\delta > 0$, we have

$$
\begin{aligned}
&\Big|\int_{\mathbb{R}} (e^{tK(z)}-1)(f(x-a_n z)-f(x))\,dz\Big| \\
&\quad\leqslant \int_{|a_n z|\leqslant\delta} (e^{tK(z)}-1)|f(x-a_n z)-f(x)|\,dz \\
&\qquad + \int_{|a_n z|>\delta} (e^{tK(z)}-1)|f(x-a_n z)-f(x)|\,dz \\
&\quad\leqslant \sup_{|a_n z|\leqslant\delta} |f(x-a_n z)-f(x)| \int_{|a_n z|\leqslant\delta} (e^{tK(z)}-1)\,dz \\
&\qquad + \int_{|a_n z|>\delta} z(e^{tK(z)}-1)\frac{f(x-a_n z)}{z}\,dz + f(x)\int_{|a_n z|>\delta} (e^{tK(z)}-1)\,dz \\
&\quad\leqslant \sup_{|y|\leqslant\delta} |f(x-y)-f(x)| \int_{\mathbb{R}} (e^{tK(z)}-1)\,dz \\
&\qquad + \frac{1}{\delta}\sup_{|a_n z|>\delta} [|z|(e^{tK(z)}-1)] \int_{|z-x|>\delta} f(z)\,dz + f(x)\int_{|a_n z|>\delta} (e^{tK(z)}-1)\,dz,
\end{aligned}
$$

which tends to 0 if one lets first $n \to \infty$, and then $\delta \to 0$. From the inequality $1 + x \leqslant e^x$ for all $x \in \mathbb{R}$, the limit (3.3) holds by a similar proof. ∎

LEMMA 3.2. *Suppose that either* **(L1)** *holds, or $K(\cdot)$ is a bounded Borel function. Then for each $x \in \mathbb{R}$ and for any $t, m > 0$, we have*

$$
\mathbb{E}e^{tK(x-mX_1)} < \infty. \tag{3.4}
$$

Proof. (1) Assume that **(L1)** is satisfied. Then

$$
\mathbb{E}(e^{tK(x-mX_1)}-1) = \frac{1}{m}\int_{\mathbb{R}} (e^{tK(u)}-1) f\left(\frac{x-u}{m}\right) du,
$$

which yields the claim (3.4) by the boundedness of f and condition (3.1).

(2) Assume that $K(\cdot)$ is a bounded Borel function. Then

$$
\int_{\mathbb{R}} (e^{tK(z)}-1) f(x-mz)\,dz = \frac{1}{m}\int_{\mathbb{R}} (e^{tK(\frac{x-u}{m})}-1) f(u)\,du \leqslant \frac{C}{m}
$$

where C is a positive constant depending on K. ∎

THEOREM 3.1. *Suppose that either* **(L1)–(L2)** *or* **(L1′)–(L2)** *hold and $K(\cdot)$ is a bounded Borel function. Then for each continuity point x of $f(\cdot)$ and for any $r > 0$, we have*

$$
\lim_{n\to\infty} \frac{1}{n^{1-\alpha}} \log \mathbb{P}(|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \geqslant r) = -J(r, x)
$$

where

$$J(r,x)=\sup_{\lambda\in\mathbb{R}}\Big\{r\lambda-f(x)\int_{\mathbb{R}}\Big[\int_0^1 s^{-\alpha}(\exp(\lambda s^{\alpha/2}K(z))-1-\lambda s^{\alpha/2}K(z))\,ds\Big]\,dz\Big\}.$$

Proof. We use the Gärtner–Ellis theorem [2, Theorem 2.3.6]. For any j, let

$$K_j=K\bigg(\frac{x-X_j}{j^{-\alpha}}\bigg).$$

Then we have (here $h_n=n^{-\alpha}$)

$$\hat f_n(x)-\mathbb{E}\hat f_n(x)=\frac{1}{n^{1-\alpha/2}}\sum_{j=1}^n j^{\alpha/2}(K_j-\mathbb{E}K_j).$$

Now we need to compute the logarithmic moment generating function of $\hat f_n(x)-\mathbb{E}\hat f_n(x)$. For any $\lambda\in\mathbb{R}$, we have

$$\begin{aligned}
&\Lambda_n(\lambda):=\frac{1}{n^{1-\alpha}}\log\mathbb{E}\exp\bigg(\frac{\lambda}{n^{\alpha/2}}\sum_{j=1}^n j^{\alpha/2}(K_j-\mathbb{E}K_j)\bigg)\\
&=\frac{1}{n^{1-\alpha}}\sum_{j=1}^n\bigg[\log\mathbb{E}\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\bigg)-\bigg(\mathbb{E}\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\bigg)-1\bigg)\bigg]\\
&\quad+\frac{1}{n^{1-\alpha}}\sum_{j=1}^n\bigg[\bigg(\mathbb{E}\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\bigg)-1\bigg)-\frac{f(x)}{j^\alpha}\int_{\mathbb{R}}\bigg(\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K(z)\bigg)-1\bigg)\,dz\bigg]\\
&\quad+\frac{1}{n^{1-\alpha}}\sum_{j=1}^n\frac{f(x)}{j^\alpha}\int_{\mathbb{R}}\bigg(\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K(z)\bigg)-1-\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K(z)\bigg)\,dz\\
&\quad+\frac{1}{n^{1-\alpha}}\sum_{j=1}^n\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}\bigg(\frac{f(x)}{j^\alpha}\int_{\mathbb{R}}K(z)\,dz-\mathbb{E}K_j\bigg)\\
&=:I_1+I_2+I_3+I_4.
\end{aligned}$$

By Lemma 3.1, for any $\varepsilon>0$ there exists a positive constant N_0 such that for $N_0\leqslant j\leqslant n$ we have

$$\begin{aligned}
(3.5)\quad&\bigg|\mathbb{E}\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\bigg)-1\bigg|\\
&=j^{-\alpha}\bigg|\int_{\mathbb{R}}\bigg(\exp\bigg(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K(z)\bigg)-1\bigg)f(x-j^{-\alpha}z)\,dz\bigg|\\
&\leqslant j^{-\alpha}\bigg|\int_{\mathbb{R}}(e^{|\lambda|K(z)}-1)f(x-j^{-\alpha}z)\,dz\bigg|\leqslant j^{-\alpha}\Big(\varepsilon+f(x)\int_{\mathbb{R}}(e^{|\lambda|K(z)}-1)\,dz\Big).
\end{aligned}$$

Now we can choose N_1 such that

$$\left|\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-1\right|\leqslant 1/2 \quad \text{for } N_1\leqslant j\leqslant n.$$

By using the elementary inequality

$$|\log(1+x)-x|\leqslant 2x^2 \quad \text{for } |x|\leqslant 1/2,$$

for $N:=\max\{N_0,N_1\}\leqslant j\leqslant n$ we have

$$\begin{aligned}&\left|\log\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-\left(\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-1\right)\right|\\ &\qquad\leqslant 2\left(\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-1\right)^2\leqslant 2j^{-2\alpha}\left(\varepsilon+f(x)\int_{\mathbb{R}}(e^{|\lambda|K(z)}-1)\,dz\right)^2.\end{aligned}$$

Furthermore, by Lemma 3.2, there exists a positive constant C depending on K, N, λ and x such that

$$\begin{aligned}I_1&=\frac{1}{n^{1-\alpha}}\sum_{j=1}^{n}\left[\log\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-\left(\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-1\right)\right]\\ &\leqslant\frac{1}{n^{1-\alpha}}\sum_{j=1}^{N}\left[\log\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-\left(\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-1\right)\right]\\ &\qquad+\frac{1}{n^{1-\alpha}}\sum_{j=N+1}^{n}\left[\log\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-\left(\mathbb{E}\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K_j\right)-1\right)\right]\\ &\leqslant\frac{CN}{n^{1-\alpha}}+\frac{2}{n^{1-\alpha}}\sum_{j=N+1}^{n}j^{-2\alpha}\left(\varepsilon+f(x)\int_{\mathbb{R}}(e^{|\lambda|K(z)}-1)\,dz\right)^2,\end{aligned}$$

which yields $I_1\to 0$. For the terms I_2 and I_4, arguing similarly we have

$$I_2=\frac{1}{n^{1-\alpha}}\sum_{j=1}^{n}j^{-\alpha}\left[\int_{\mathbb{R}}\left(\exp\left(\frac{\lambda j^{\alpha/2}}{n^{\alpha/2}}K(z)\right)-1\right)(f(x-j^{-\alpha}z)-f(x))\,dz\right]\to 0$$

and

$$I_4=\frac{\lambda}{n^{1-\alpha/2}}\sum_{j=1}^{n}j^{-\alpha/2}\left(f(x)\int_{\mathbb{R}}K(z)\,dz-\int_{\mathbb{R}}K(z)f(x-j^{-\alpha}z)\,dz\right)\to 0.$$

For the term I_3, by the elementary inequality

$$|e^x-1-x|\leqslant\frac{x^2}{2}e^{|x|} \quad \text{for all } x\in\mathbb{R},$$

we have

$$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{n}{j}\right)^{\alpha}\left|\exp\left(\lambda\left(\frac{j}{n}\right)^{\alpha/2}K(z)\right)-1-\lambda\left(\frac{j}{n}\right)^{\alpha/2}K(z)\right|\leqslant\frac{\lambda^2}{2}K^2(z).$$

Since **(L1)** or **(L1′)** implies $\int_{\mathbb{R}} K^2(z)\,dz<\infty$, from some analytic considerations and the dominated convergence theorem we have

$$\begin{aligned}I_3 &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{n}{j}\right)^{\alpha}\int_{\mathbb{R}}\left(\exp\left(\lambda\left(\frac{j}{n}\right)^{\alpha/2}K(z)\right)-1-\lambda\left(\frac{j}{n}\right)^{\alpha/2}K(z)\right)f(x)\,dz\\ &\to f(x)\int_{\mathbb{R}}\left[\int_0^1 s^{-\alpha}\big(\exp(\lambda s^{\alpha/2}K(z))-1-\lambda s^{\alpha/2}K(z)\big)\,ds\right]dz.\end{aligned}$$

From the above discussion, the desired result follows. ∎

COROLLARY 3.1. *Under the conditions* **(L1′)–(L2)**, *assume that K is bounded and has Fourier transform $K^*(u)=\int_{-\infty}^{\infty}e^{-iuy}K(y)\,dy$. Suppose further that for some positive constant β,*

$$\lim_{u\to 0}\,[1-K^*(u)]/|u|^{\beta}=k_{\beta}$$

is finite and

$$f^{(\beta)}(x)=-\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-iux}|u|^{\beta}\phi(u)\,du$$

exists, where $\phi(\cdot)$ denotes the characteristic function of the random variable X with density function $f(x)$. Then for any α with $\alpha(\beta+2^{-1})<1$ and any $r>0$, we have

$$\lim_{n\to\infty}\frac{1}{n^{1-\alpha}}\log\mathbb{P}\big(|\hat{f}_n(x)-(1-\alpha/2)^{-1}f(x)|\geqslant r\big)=-J(r,x)$$

where $J(r,x)$ is defined in Theorem 3.1.

Proof. Note that the condition **(L1′)** implies

$$\int_{\mathbb{R}}K(z)\,dz<\infty\quad\text{and}\quad\lim_{|z|\to\infty}|z|K(z)=0.$$

Then the corollary can be deduced from Theorems 2.1 and 3.1. ∎

Acknowledgments. The authors are grateful to the referees for their valuable reports which improved the presentation of this work.

REFERENCES

[1] M. Csörgö and L. Horváth, *Central limit theorems for L_p-norms of density estimators,* Probab. Theory Related Fields 80 (1988), 269–291.

[2] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications,* Springer, Berlin, 2010.

[3] F. Q. Gao, *Moderate deviations and large deviations for kernel density estimators,* J. Theoret. Probab. 16 (2003), 401–418.

[4] F. Q. Gao, *Moderate deviations and law of the iterated logarithm in $L_1(\mathbb{R}^d)$ for kernel density estimators,* Stochastic Process. Appl. 118 (2008), 452–473.

[5] H. Y. Liang and J. Baek, *Asymptotic normality of recursive density estimates under some dependence assumptions,* Metrika 60 (2004), 155–166.

[6] H. Y. Liang and J. Baek, *Berry–Esseen bounds for density estimates under NA assumption,* Metrika 68 (2008), 305–322.

[7] D. Louani, *Large deviations limit theorems for the kernel density estimator,* Scand. J. Statist. 25 (1998), 243–253.

[8] D. Louani, *Large deviations for the L_1-distance in kernel density estimation,* J. Statist. Plann. Inference 90 (2000), 177–182.

[9] Z. D. Lu, *Asymptotic normality of kernel density estimators under dependence,* Ann. Inst. Statist. Math. 53 (2001), 447–468.

[10] E. Masry, *Recursive probability density estimation for weakly dependent stationary processes,* IEEE Trans. Inform. Theory 32 (1986), 254–267.

[11] E. Masry and L. Györfi, *Strong consistency and rates for recursive probability density estimators of stationary processes,* J. Multivariate Anal. 22 (1987), 79–93.

[12] A. Mokkadem, M. Pelletier and B. Thiam, *Large and moderate deviations principles for recursive kernel estimator of a multivariate density and its partial derivatives,* Serdica Math. J. 32 (2006), 323–354.

[13] A. Mokkadem, M. Pelletier and J. Worms, *Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives,* Austral. New Zealand J. Statist. 47 (2005), 489–502.

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

[15] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function,* Ann. Math. Statist. 27 (1956), 832–837.

[16] E. J. Wegman and H. I. Davies, *Remarks on some recursive estimators of a probability density,* Ann. Statist. 7 (1979), 316–327.

[17] C. T. Wolverton and T. J. Wagner, *Asymptotically optimal discriminant functions for pattern classification,* IEEE Trans. Information Theory 15 (1969), 258–265.

[18] H. Yamato, *Sequential estimation of a continuous probability density function and mode,* Bull. Math. Statist. 14 (1971), 1–12.

[19] D. X. Zhang and H. Y. Liang, *Recursive density estimation of NA samples,* Chinese J. Appl. Probab. Statist. 24 (2008), 123–134.

Yu Miao, Qinghui Gao, Jianyong Mu, Conghui Deng
College of Mathematics and Information Science
Henan Normal University
Xinxiang, Henan Province, 453007, China
E-mail: yumiao728@gmail.com, qinghuigaocb@163.com,
jianyongmu@163.com, conghuideng820@163.com

Received 28.2.2018;
revised version 7.10.2018