Available online at www.sciencedirect.com
SCIENCE DIRECT®
Stochastic Processes and their Applications 115 (2005) 275–298
stochastic processes and their applications
www.elsevier.com/locate/spa

Large deviations of kernel density estimator in $L^1(\mathbb{R}^d)$ for uniformly ergodic Markov processes ☆

Liangzhen Lei[a,b,*], Liming Wu[a,b]

[a]Department of Mathematics, Wuhan University, 430072 Hubei, PR China
[b]Laboratoire de Mathématiques, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France

Received 24 October 2003; received in revised form 17 May 2004; accepted 15 September 2004
Available online 8 October 2004

Abstract

In this paper, we consider a uniformly ergodic Markov process $(X_n)_{n\geqslant 0}$ valued in a measurable subset E of $\mathbb{R}^d$ with the unique invariant measure $\mu(\mathrm{d}x)=f(x)\,\mathrm{d}x$, where the density f is unknown. We establish the large deviation estimations for the nonparametric kernel density estimator f_n^* in $L^1(\mathbb{R}^d,\mathrm{d}x)$ and for $\|f_n^*-f\|_{L^1(\mathbb{R}^d,\mathrm{d}x)}$, and the asymptotic optimality f_n^* in the Bahadur sense. These generalize the known results in the i.i.d. case.

Keywords: Large deviations; Kernel density estimator; Donsker–Varadhan entropy; Uniformly ergodic Markov process; Bahadur efficiency

1. Introduction

Let $\{X_n; n\geqslant 0\}$ be a Doeblin recurrent Markov chain valued in a Borel measurable subset E of $\mathbb{R}^d$, defined on the probability space $(\Omega,(\mathscr{F}_n)_{n\in\mathbb{N}},\mathscr{F},(\mathbb{P}_x)_{x\in E})$, with (unknown) transition kernel $P(x,\mathrm{d}y)$. Moreover, we assume that the unique

☆Research supported by the Yangtze professorship program.
*Corresponding author. Laboratoire de Mathématiques, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France. Tel.: +33 04 73 40 53 92; fax: +33 04 73 40 70 64.
E-mail address: lei@math.univ-bpclermont.fr (L. Lei).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2004.09.004

invariant measure μ of P is absolutely continuous, i.e., $\mu(\mathrm{d}x) = f(x)\,\mathrm{d}x$ where the density f is unknown.

Let K be a measurable function such that

$$K \geqslant 0, \quad \int_{\mathbb{R}^d} K(x)\,\mathrm{d}x = 1 \tag{1.1}$$

and set $K_h(x) = \frac{1}{h^d} K\left(\frac{x}{h}\right)$. Given the observed sample $\{X_0, \ldots, X_{n-1}\}$, we consider the empirical measure $L_n = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{X_i}$ and define the kernel density estimator of the unknown f as usually as

$$f_n^*(x) = K_{h_n} * \mathrm{d}L_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d, \tag{1.2}$$

where $\{h_n, n \geqslant 0\}$ is a sequence of positive numbers (bandwidth) satisfying

$$h_n \to 0, \quad nh_n^d \to +\infty \quad \text{as } n \to \infty. \tag{1.3}$$

A natural measure of closeness of f_n^* to the unknown f is its $L^1(\mathbb{R}^d) := L^1(\mathbb{R}^d, \mathrm{d}x)$ distance below,

$$D_n^* = \int_{\mathbb{R}^d} |f_n^*(x) - f(x)|\,\mathrm{d}x. \tag{1.4}$$

The limit behavior of f_n^* in $L^1(\mathbb{R}^d)$ is a subject of current study.

In the i.i.d. case, Devroye [6] proved that all types of $L^1(\mathbb{R}^d)$-consistency are equivalent to condition (1.3) on the bandwidth (h_n). Csörgö and Horváth [3] and Horváth [11] investigated the asymptotic normality of D_n^*. Louani [16] established the large deviation principle (LDP in short) of D_n^*. Gao [8] obtained the LDP and the moderate deviation principle of f_n^* in $L^\infty(\mathbb{R}^d)$. And recently Lei et al. [14] prove the weak LDP of f_n^* in $L^1(\mathbb{R}^d)$, and show that the corresponding LDP is false. More recently Gao [9] obtains the moderate deviation principle of f_n^* in $L^1(\mathbb{R}^d)$ and the law of the iterated logarithm for D_n^*. Giné et al. [10] establish a functional central limit theorem and a Glivenko–Cantelli theorem.

How to extend those results from the i.i.d. case to Markov processes (or dependent case) is a very natural and important question. In fact, numerous practical models from economic time series or biologies are Markov process (cf. [2]), for which it is very important to estimate the asymptotic equilibrium measure $\mu(\mathrm{d}x) = f(x)\,\mathrm{d}x$. Known works in the dependent case are concentrated on the consistency of f_n^* and its asymptotic normality, see Peligrad [18], Bosq et al. [1] and the references therein. But few are known about the large deviations of f_n^* and D_n^* in the dependent case.

In a recent work [15], as a first step towards the large deviations of f_n^*, we prove the exponential convergence of f_n^* to f for a ϕ-mixing sequence (X_n). In this paper which is a sequel to [15], we investigate the large deviations of f_n^* in $L^1(\mathbb{R}^d)$ and of D_n^* in the framework of uniformly ergodic Markov chains (see H1 below).

Large deviation of occupation measures L_n for Markov processes is a traditional subject in probability, initiated by Donsker and Varadhan [7]. The rate function is the Donsker–Varadhan level-2 entropy given by

$$J(\nu) := \sup\left\{ \int \log \frac{u}{Pu}\, d\nu;\ 1 \leqslant u \in b\mathscr{B}(E) \right\}, \quad \forall \nu \in M_1(E), \tag{1.5}$$

where $b\mathscr{B}(E)$ is the space of real bounded functions measurable w.r.t. the Borel σ-field $\mathscr{B}(E)$ of E, and $M_1(E)$ denotes the space of all probability measures on E.

Deuschel and Stroock [5, Theorem 4.1.14] obtained the LDP of L_n w.r.t. the τ-topology (i.e., the weakest topology on $M_1(E)$ such that $\nu \to \nu(f) := \int_E f(x)\, d\nu(x)$ is continuous for all $f \in b\mathscr{B}(E)$), under the following:

H1 (*Uniform ergodicity*). There are $1 \leqslant l \leqslant N \in \mathbb{N}$ and $M \geqslant 1$ such that

$$P^l(x, A) \leqslant M \frac{P(y, A) + \cdots + P^N(y, A)}{N}, \quad \forall x, y \in E, A \in \mathscr{B}(E).$$

Later, a lot of significant progress has been made, see [4,23,13] and the references therein.

This paper is organized as follows. The main results such as the weak*-LDP of f_n^* on $L^1(\mathbb{R}^d)$, the large deviation estimation for $\mathbb{P}_x(D_n^* > \delta)$ and the asymptotic efficiency of the estimator f_n^* in the Bahadur sense, etc. are presented in the next section. Those results are, as far as we know, obtained for the first time in the dependent case. In Section 3, we prepare several lemmas. We give the proofs of the main results in Sections 4–7.

2. Main results

Throughout this paper, we adopt the following notations. $L^p(\mathbb{R}^d) := L^p(\mathbb{R}^d, dx)$, $L^p(\mu) := L^p(E, \mu)$; $\|f\|_1 = \|f\|_{L^1(\mathbb{R}^d, dx)}$. We denote by $b\mathscr{B}$ (resp. $b\mathscr{B}(E)$) the space of all real bounded and Borel $\mathscr{B}$-measurable functions on $\mathbb{R}^d$ (resp. E) equipped with the sup norm $\|V\| = \sup_x |V(x)|$. We write $\nu(V) = \langle V \rangle_\nu := \int_E V(x)\, d\nu(x)$. Without loss of generality, we assume that $(X_n)_{n \geqslant 0}$ is the system of coordinates on $\Omega := E^{\mathbb{N}}$ and $\mathbb{P}_x$ is the law of the Markov chain with the transition kernel P and the starting point $x \in E$. Set $\mathbb{P}_\nu(\cdot) := \int_E \mathbb{P}_x(\cdot)\, d\nu(x)$ and $\mathbb{E}^\nu(\cdot) = \int_\Omega \cdot\, d\mathbb{P}_\nu$. Let $(\theta\omega)_n := \omega_{n+1}$ ($n \in \mathbb{N}$) be the shift on Ω.

When the bandwidth $h_n \to 0$, $f_n^*\, dx$ is "close" to L_n in the τ-topology, so we may hope that $f_n^*\, dx$ satisfies the same LDP as L_n. This intuition is true:

Theorem 2.1. *Assume* H1 *and* $h_n \to 0$ (*without* (1.3)). *Then* $\mathbb{P}_x(f_n^* \in \cdot)$ *satisfies, uniformly for the initial points* $x \in E$, *the LDP in* $L^1(\mathbb{R}^d)$ *w.r.t. the weak topology* $\sigma(L^1, L^\infty)$, *with the rate function given by*

$$J(g) := \begin{cases} J(g\, dx) & \text{if } g \in \mathscr{P}(E), \\ +\infty & \text{if } g \in L^1(\mathbb{R}^d) \backslash \mathscr{P}(E). \end{cases} \tag{2.1}$$

Here $J(\cdot)$ is the Donsker–Varadhan level-2 entropy given in (1.5), *$\mathscr{P}(E)$ is the set of all probability density functions on $\mathbb{R}^d$ with support in E, i.e., those $g \in L^1(\mathbb{R}^d)$ such that $g \geqslant 0$ on $\mathbb{R}^d$, $g = 0$, a.e. on $E^c := \mathbb{R}^d \backslash E$ and $\int_{\mathbb{R}^d} g \, dx = 1$.*

More precisely, J is inf-compact on $(L^1(\mathbb{R}^d), \sigma(L^1, L^\infty))$, and for any measurable subset A of $L^1(\mathbb{R}^d)$,

$$
\begin{aligned}
- \inf_{g \in A^{o\sigma}} J(g) &\leqslant \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x \in E} \mathbb{P}_x(f_n^* \in A) \\
&\leqslant \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x \in E} \mathbb{P}_x(f_n^* \in A) \leqslant - \inf_{g \in \bar{A}^\sigma} J(g),
\end{aligned}
$$

where $A^{o\sigma}, \bar{A}^\sigma$ denote, respectively, the interior and the closure of A w.r.t. the weak topology $\sigma(L^1, L^\infty)$.

The LDP w.r.t. the weak topology on $L^1(\mathbb{R}^d)$ above is of the same type as the classical results for L_n w.r.t. the τ-topology. But it is too weak in the sense that it does not entail the consistency, i.e., $D_n^* \to 0$ in probability. For statistical issues, the main objects to be studied are

(i) $\mathbb{P}_x(\|f_n^* - g\|_1 < \delta)$ where $g \in \mathscr{P}(E)$ is fixed, which is important in the hypothesis testing: $H_0 : d\mu(x) = f(x)\,dx$ against $H_1 : d\mu(x) = g(x)\,dx$; or
(ii) $\mathbb{P}_x(D_n^* > \delta)$, whose statistical importance is obvious.

Unfortunately Theorem 2.1 cannot be applied for them, since $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - g\|_1 < \delta\}$ is not open in $\sigma(L^1, L^\infty)$ and $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - f\|_1 \geqslant \delta\}$ is not closed in $\sigma(L^1, L^\infty)$. They are objects of

Theorem 2.2. *Assume* H1 *and* (1.3). *Then $\mathbb{P}_x(f_n^* \in \cdot)$ satisfies, uniformly for initial state $x \in E$, the weak*-LDP on $(L^1(\mathbb{R}^d), \|\cdot\|_1)$ with the rate function $J(g)$ given by* (2.1), *i.e., for any $g \in L^1(\mathbb{R}^d)$,*

$$
\begin{aligned}
&\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x \in \mathbb{R}^d} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \\
&\quad = \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x \in \mathbb{R}^d} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) = -J(g).
\end{aligned}
\tag{2.2}
$$

Notice that the corresponding (good) LDP is in general not true, because even in the i.i.d. case, $J(g) = J^{iid}(g) = \int g(x) \log \frac{g(x)}{f(x)} \, dx$ (for $g \in \mathscr{P}(E)$ and $g\,dx \ll f\,dx$) is not inf-compact on $(L^1(\mathbb{R}^d), \|\cdot\|_1)$ (as noted in [14]).

Theorem 2.3. *Assume* H1 *and* (1.3). *Then*

(a) *For any $\delta > 0$,*

$$
\begin{aligned}
-I(\delta) &\leqslant \liminf_{n\to\infty} \frac{1}{n} \log \inf_{x \in \mathbb{R}^d} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \\
&\leqslant \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x \in \mathbb{R}^d} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \leqslant - I(\delta-),
\end{aligned}
\tag{2.3}
$$

where

$$I(\delta) = \inf\{J(g);\ g \in \mathscr{P}(E), \|g-f\|_1 > \delta\}. \tag{2.4}$$

(b) *We have for any* $\delta > 0$,

$$I(\delta) \geqslant \frac{1}{l}\,(I^{iid}(\delta) - \log M), \tag{2.5}$$

where l, M *are given in* H1 *and* $I^{iid}(\delta)$ *is the rate function of the LDP of* $\|f_n^* - f\|_1$ *in the case where* (X_n) *are i.i.d. of common law* μ (*see* (2.9) *below*).

(c) *Besides* H1, *assume that* P *is aperiodic. Then we also have*

$$I(\delta) \geqslant \frac{\delta^2}{8(1+S)^2}, \quad \forall \delta > 0, \tag{2.6}$$

where $S := \sum_{k=1}^{\infty} \sup_{x,y \in E} \|P^k(x,\cdot) - P^k(y,\cdot)\|_{\mathrm{TV}}$ (*here* $\|\cdot\|_{\mathrm{TV}}$ *denotes the total variation*) *is finite.*

Remark 2.1. Parts (b) and (c) of Theorem 2.3 are served for δ large or small, respectively. By the contraction principle and the LDP of L_n under H1 in [5, Theorem 4.1.14], for each $V \in b\mathscr{B}(E)$, $L_n(V) - \mu(V)$ satisfies the LDP with the inf-compact rate function given by

$$J_V(r) = \inf\{J(v);\ v(V) = \mu(V) + r\}, \quad \forall r \in \mathbb{R}. \tag{2.7}$$

Since $J_V(0) = 0$ and J_V is convex with values in $[0, +\infty]$, J_V is non-decreasing and left continuous on $[0, +\infty)$. Consequently using $\|v - \mu\|_{\mathrm{TV}} = \sup_{\|V\| \leqslant 1}[v(V) - \mu(V)] = 2\sup_{A \in \mathscr{B}} |v(A) - \mu(A)|$ (for two probability measures μ, v), we can identify $I(\delta)$ given in (2.4) as

$$\begin{aligned} I(\delta) &= \inf\{J(v)|\ \sup_{\|V\| \leqslant 1}[v(V) - \mu(V)] > \delta\} \\ &= \inf_{\|V\| \leqslant 1} \inf_{r > \delta} J_V(r) = \inf_{\|V\| \leqslant 1} J_V(\delta+) \\ &= \inf\left\{ J(v)|\ \sup_{A \in \mathscr{B}(E)} [(v(A) - \mu(A)] > \delta/2 \right\} = \inf_{A \in \mathscr{B}(E)} J_A(\delta/2+), \end{aligned} \tag{2.8}$$

where $J_A = J_{1_A}$. In the i.i.d. case, the last expression in (2.8) above coincides exactly with the rate function of the LDP for D_n^* found by Louani [16]. Indeed, when $\mu(A) = a \in (0,1)$, then for any $\delta > 0$, $J_A^{iid}(\delta/2)$ is given by

$$\Gamma_a^+(\delta) = \begin{cases} \left(a + \frac{\delta}{2}\right)\log\left(1 + \frac{\delta}{2a}\right) + \left(1 - a - \frac{\delta}{2}\right)\log\left(1 - \frac{\delta}{2(1-a)}\right) & \text{if } 0 < \delta < 2 - 2a, \\ +\infty, & \text{otherwise} \end{cases}$$

(then $J_A^{iid}(\delta/2) = J_A^{iid}(\delta/2+)$) and

$$I^{iid}(\delta) := \inf_{a \in (0,1)} \Gamma_a^+(\delta) = \inf_A J_A^{iid}(\delta/2) \tag{2.9}$$

which is I^{iid} in [16].

Remark 2.2. If I were strict increasing on $(0,a)$ where $a:=\sup\{r>0; I(r)<+\infty\}$, then we can prove in fact the LDP of D_n^* in $\mathbb{R}^+$ with the rate function $\delta\to I(\delta-)$, from (2.3).

In the results above, we have the large deviation estimates of the estimator f_n^*, useful in statistics. We now show that f_n^* is asymptotically optimal in the Bahadur sense. Let Θ be the set of unknown data (P,μ) verifying H1 and $\mu(\mathrm{d}x)\ll \mathrm{d}x$. Given a subset $\mathscr{D}$ of the unit ball in $b\mathscr{B}$, we say that an estimator $T_n(\cdot):=T_n(\cdot;X_0,\dots,X_{n-1})\in L^1(\mathbb{R}^d)$ is an asymptotically $\sigma(L^1,\mathscr{D})$-consistent estimator of the density f, if $\forall V\in\mathscr{D}$,

$$\int_{\mathbb{R}^d} T_n(x)V(x)\,\mathrm{d}x \to \int_{\mathbb{R}^d} f(x)V(x)\,\mathrm{d}x$$

in probability measure $\mathbb{P}_\mu$. From the results above, we shall derive:

Theorem 2.4. *Given $(P,\mu)\in\Theta$, let $((X_n),(\mathbb{P}_x)_{x\in E})$ be the associated Markov process.*

(a) (Bahadur type lower bound). *Assume that $\mathscr{D}$ is dense in the unit ball of $L^\infty(\mathbb{R}^d)$ w.r.t. the weak* topology $\sigma(L^\infty,L^1)$. Then for any $\sigma(L^1,\mathscr{D})$-asymptotically consistent estimator T_n of the density f,*

$$\liminf_{r\to 0+}\frac{1}{r^2}\liminf_{n\to\infty}\frac{1}{n}\log \mathbb{P}_\mu(\|T_n-f\|_1>r)$$
$$\geqslant -\frac{1}{2\sup_{\|V\|\leqslant 1}\sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathscr{B}}\sigma^2(1_A)},\tag{2.10}$$

where

$$\sigma^2(V):=\mathrm{Var}_\mu(V)+2\sum_{k=1}^{\infty}\langle V-\mu(V),P^kV\rangle_\mu.$$

If moreover $\|T_n-T_n\circ\theta^N\|_1\leqslant\delta_n\to 0$, then (2.10) *still holds with $\mathbb{P}_\mu$ substituted by $\inf_{x\in E}\mathbb{P}_x$.*

(b) (Asymptotic efficiency of f_n^* in the Bahadur sense). *If h_n verifies* (1.3), *then*

$$\liminf_{r\to 0+}\frac{1}{r^2}\lim_{n\to\infty}\frac{1}{n}\log\inf_{x\in E}\mathbb{P}_x(\|f_n^*-f\|_1>r)$$
$$=\limsup_{r\to 0+}\frac{1}{r^2}\lim_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x(\|f_n^*-f\|_1>r)$$
$$=-\frac{1}{2\sup_{\|V\|\leqslant 1}\sigma^2(V)}=-\frac{1}{8\sup_{A\in\mathscr{B}}\sigma^2(1_A)}.\tag{2.11}$$

Thus f_n^* is an asymptotically efficient estimator of f in the Bahadur sense. And $1/\sigma^2(V)$ can be interpreted as the Fisher information at the direction V of our statistical model Θ.

All the results above except perhaps Theorem 2.4(a) are, as far as we know, new in the dependent case.

Remark 2.3. In comparison with the i.i.d. case, the new object in the Markov chain case is the transition kernel density $p(x,y) := P(x, \mathrm{d}y)/\mathrm{d}y$. For its estimation or more precisely $F(x,y) := f(x)p(x,y)$, no more effort is required due to the subtleness of our assumption H1. Indeed, consider the Markov chain $Y_n := (X_n, X_{n+1})$ with values in E^2, whose transition kernel still verifies H1 and whose unique invariant measure is $F(x,y)\,\mathrm{d}x\,\mathrm{d}y$. The Donsker–Varadhan level-2 entropy for this new Markov chain possesses an explicit expression [5]:

$$J^{(2)}(Q) := \begin{cases} \int\int_{E\times E} Q(\mathrm{d}x,\mathrm{d}y)\log\frac{Q(x,\mathrm{d}y)}{P(x,\mathrm{d}y)} & \text{if } Q \in M_1^s(E^2),\ Q(x,\cdot) \ll P(x,\cdot), \\ +\infty & \text{otherwise,} \end{cases} \tag{2.12}$$

where $Q \in M_1^s(E^2)$ iff $Q \in M_1(E^2)$ and $Q(A \times E) = Q(E \times A)$, $\forall A \in \mathscr{B}(E)$, and $Q(x, \mathrm{d}y)$ is the regular conditional distribution of the second coordinate X_1 knowing the first $X_0 = x$. Consider the kernel density estimator

$$F_n^*(x,y) := \frac{1}{n}\sum_{k=0}^{n-1} K_{h_n}(x - X_k)\cdot K_{h_n}(y - X_{k+1}).$$

Hence the previous results apply for F_n^* if condition (1.3) is substituted by $h_n \to 0$ and $nh_n^{2d} \to +\infty$.

3. Several lemmas

For every $V \in b\mathscr{B}(E)$, put $P^V(x,\mathrm{d}y) := \mathrm{e}^{V(x)}P(x,\mathrm{d}y)$. We have the Feynman–Kac formula

$$(P^V)^n f(x) = \mathbb{E}^x f(X_n)\exp\sum_{k=0}^{n-1} V(X_k).$$

Let $\|(P^V)^n\| := \sup_{\|f\|\leqslant 1}\|(P^V)^n f\| = \|(P^V)^n 1\|$ be the norm of P^V acting on $b\mathscr{B}(E)$. Consider the uniform Cramer functional [5]

$$\Lambda(V) = \lim_{n\to\infty}\frac{1}{n}\log\|(P^V)^n\| = \lim_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{E}^x\exp\left(\sum_{k=0}^{n-1}V(X_k)\right),$$

then $\mathrm{e}^{\Lambda(V)}$ is the spectral radius of P^V on $b\mathscr{B}(E)$. It is well known [5] that

$$J(\nu) = \sup\{\nu(V) - \Lambda(V);\ V \in b\mathscr{B}(E)\}, \quad \forall \nu \in M_1(E). \tag{3.1}$$

By the LDP of L_n in [5] and the Laplace principle due to Varadhan, $\forall V \in b\mathscr{B}(E)$,

$$\Lambda(V) = \sup\{\nu(V) - J(\nu);\ \nu \in M_1(E)\} = \sup\left\{\int gV\,\mathrm{d}\mu - J(g);\ g \in \mathscr{P}(E)\right\}, \tag{3.2}$$

where the second equality follows from the fact that if $J(\nu) < +\infty$, then $\nu \ll \mu$ under H1 (see [23, B.23]).

By H1, $P^l(x, \mathrm{d}y) \leqslant M\mu(\mathrm{d}y)$. Hence for each $V \in b\mathscr{B}(E)$,

$$\Lambda(V) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}^{\mu} \exp\left(\sum_{k=0}^{n-1} V(X_k)\right). \tag{3.3}$$

Lemma 3.1. *For positive operator P^V defined as above, let $(P^V)^*$ be the dual operator of P^V w.r.t. μ. Then*

(a) *There exist $\phi \in b\mathscr{B}(E)$, $\psi \in b\mathscr{B}(E)$ both strictly positive, such that*

$$P^V\phi = \mathrm{e}^{\Lambda(V)}\phi \ \text{ over } E, \quad (P^V)^*\psi = \mathrm{e}^{\Lambda(V)}\psi,\ \mu\text{-a.s.}$$

and the following Harnack inequalities hold:

$$\frac{\phi(y)}{\phi(x)} \vee \frac{\psi(y)}{\psi(x)} \leqslant \frac{M}{N} \cdot \mathrm{e}^{2N\|V\|} \cdot \frac{\sum_{k=1}^{N} \mathrm{e}^{k\Lambda(V)}}{\mathrm{e}^{l\Lambda(V)}} \leqslant M\mathrm{e}^{3N\|V\|}, \quad \forall x, y \in E. \tag{3.4}$$

(b) *Put*

$$Q^V(x, \mathrm{d}y) = \frac{\phi(y)}{\mathrm{e}^{\Lambda(V)}\phi(x)}\, \mathrm{e}^{V(x)} P(x, \mathrm{d}y),$$

then Q^V is Doeblin recurrent, and $\nu_V := \phi\psi\mu$ is the unique invariant probability measure for Q^V.

Proof. (a) Under H1, $P^l(x, \mathrm{d}y) \leqslant M\mu(\mathrm{d}y)$ and then $P^N(x, \mathrm{d}y) \leqslant M\mu(\mathrm{d}y)$. Thus $(P^V)^N$ is uniformly integrable in $L^\infty(\mu)$ in the terms of [23]. By Theorem 3.2 in [23], there exists some $0 \leqslant \varphi \in L^\infty(\mu)$ such that $\mu(\varphi) > 0$ and

$$(P^V)^N\varphi = r^N\varphi, \quad \mu\text{-a.s.},$$

where r is the spectral radius of P^V in $L^\infty(\mu)$. Since $(P^V)^N(x, \mathrm{d}y) \leqslant \mathrm{e}^{N\|V\|}M\mu(\mathrm{d}y)$, then letting $g := (P^V)^N\varphi$, we see that $(P^V)^N g = r^N g$ everywhere over E. By (3.3), $r = \mathrm{e}^{\Lambda(V)}$. Finally setting

$$\phi(x) = \sum_{k=1}^{N} (P^V)^k g(x),$$

which is strictly positive by H1, we have for all $x \in E$,

$$P^V\phi(x) = r\phi(x) = \mathrm{e}^{\Lambda(V)}\phi(x), \quad \forall x \in E.$$

Since for any x, y,

$$\frac{\phi(y)}{\phi(x)} = \frac{(P^V)^l\phi(y)}{\sum_{k=1}^{N}(P^V)^k\phi(x)} \cdot \frac{\sum_{k=1}^{N}\mathrm{e}^{k\Lambda(V)}}{\mathrm{e}^{l\Lambda(V)}},$$

using H1 and $-\|V\|\leqslant V(x)\leqslant\|V\|$, we get

$$\frac{\phi(y)}{\phi(x)}\leqslant\frac{M}{N}\cdot\mathrm{e}^{2N\|V\|}\cdot\frac{\sum_{k=1}^{N}\mathrm{e}^{k\Lambda(V)}}{\mathrm{e}^{l\Lambda(V)}}$$

where the desired Harnack inequality (3.4) for ϕ follows.

For the corresponding result about $(P^V)^*$, we choose a kernel $P^*(x,\mathrm{d}y)$, which is the dual of P (w.r.t. μ) and also satisfies H1. Applying the previous argument to $\mathrm{e}^{V(y)}P^*(x,\mathrm{d}y)$ which is the dual of P^V (w.r.t. μ), we get the existence of ψ and the Harnack inequality (3.4) for ψ.

(b) It is easy to verify that Q^V is a Markov kernel, and $\phi\psi\mu$ is an invariant measure of Q^V. As Q^V again satisfies H1 by part (a), it is Doeblin recurrent. Then $\phi\psi\mu$ is the unique invariant measure of Q^V. $\square$

Lemma 3.2. *Under* H1, *we have for every* $V\in b\mathscr{B}(E)$ *such that* $\|V\|\leqslant 1$, $\forall r>0$, $n\geqslant 1$ *so that* $4N/n\leqslant r$,

$$\sup_{x\in E}\mathbb{P}_x\left(\frac{1}{n}\sum_{k=0}^{n-1}V(X_k)>\mu(V)+r\right)\leqslant M\exp\left(-nJ_V\left(r-\frac{4N}{n}\right)\right),\tag{3.5}$$

where $J_V(r)$ *is the rate function governing the LDP of* $L_n(V)-\mu(V)$, *given in* (2.7).

Notice that in the i.i.d. case, $M=N=1$ and (3.5) is exactly the well-known Cramer inequality. This lemma is basic to Theorem 2.3.

Proof (*following closely [5]*). (1) At first by Deuschel and Stroock [5, Lemma 4.1.4],

$$p_n(r):=\inf_{x\in E}\mathbb{P}_x\left(\frac{1}{n}\sum_{k=0}^{n-1}V(X_k)>\mu(V)+r\right)$$

is super-multiplicative, i.e., $p_{n+m}\geqslant p_np_m$, $\forall n,m\in\mathbb{N}^*$. Thus

$$\frac{1}{n}\log p_n(r)\leqslant\sup_{m\geqslant 1}\frac{\log p_m(r)}{m}=\lim_{m\to\infty}\frac{\log p_m(r)}{m}.$$

But by the uniform LDP of $L_n(V)=\frac{1}{n}\sum_{k=0}^{n-1}V(X_k)$ in [5] and the increasingness of J_V on $\mathbb{R}^+$, we have $\lim_{m\to\infty}\frac{\log p_m(r)}{m}\leqslant -J_V(r)$ for every $r\geqslant 0$. Thus

$$\inf_{x\in E}\mathbb{P}_x\left(\frac{1}{n}\sum_{k=0}^{n-1}V(X_k)>\mu(V)+r\right)\leqslant\mathrm{e}^{-nJ_V(r)},\quad\forall n\geqslant 1, r\geqslant 0.\tag{3.6}$$

(2) For every $k=1,\dots,N$, since

$$|L_n(V)\circ\theta^k-L_n(V)|\leqslant\frac{2k}{n}\leqslant\frac{2N}{n},$$

letting $\varepsilon = \frac{2N}{n}$, we have for any $r \in \mathbb{R}$, $n \geqslant 1$ and $x \in E$,

$$\begin{aligned} f_{n,r}(x) &:= \mathbb{P}_x(L_n(V) > \mu(V) + r) \leqslant \mathbb{P}_x(L_n(V) \circ \theta^k > \mu(V) + r - \varepsilon) \\ &= (P^k f_{n,r-\varepsilon})(x) \end{aligned}$$

and similarly

$$f_{n,r}(x) \geqslant \mathbb{P}_x(L_n(V) \circ \theta^k > \mu(V) + r + \varepsilon) = (P^k f_{n,r+\varepsilon})(x).$$

Thus using H1, we obtain for any $x, y \in E$,

$$f_{n,r}(x) \leqslant (P^l f_{n,r-\varepsilon})(x) \leqslant M \frac{1}{N} \sum_{k=1}^{N} (P^k f_{n,r-\varepsilon})(y) \leqslant M f_{n,r-2\varepsilon}(y).$$

Hence the desired result follows by (3.6). □

The following result is technically crucial for all results in this paper.

Lemma 3.3. (a) *$\Lambda(V)$ is Gateaux-differentiable on $b\mathscr{B}(E)$.*
(b) *If $V_n \to V$ in measure μ and $\sup_n \|V_n\| \leqslant C$, then $\Lambda(V_n) \to \Lambda(V)$.*

Proof. (a) Under H1, $(P^V)^N$ is uniformly integrable in $L^\infty(\mu)$, then by [23, Proposition 2.1], $(P^V)^{2N}$ is compact in $L^\infty(\mu)$. Consequently by the perturbation theory of linear operators [12, Chapter VII, Theorem 1.8], the largest eigenvalue $e^{2N\Lambda(V)}$ of $(P^V)^{2N}$, is real-analytic, i.e., $\Lambda(V + t\tilde{V})$ is analytic on $t \in \mathbb{R}$ for any $V, \tilde{V} \in b\mathscr{B}$ fixed.
(b) At first $\liminf_{n\to\infty} \Lambda(V_n) \geqslant \Lambda(V)$ by (3.2). Notice that $e^{N\Lambda(V)}$ is the spectral radius of $(P^V)^N$ in $L^\infty(\mu)$. Now the inverse inequality $\limsup_{n\to\infty} \Lambda(V_n) \leqslant \Lambda(V)$, follows by [23, Proposition 3.8] applied to $\pi_n := (P^{V_n})^N$. □

Lemma 3.4 (*Gibbs type principle*). *Given a function $V \in b\mathscr{B}(E)$, a probability measure ν on E satisfies*

$$J(\nu) = \langle \nu, V \rangle - \Lambda(V)$$

iff $\nu = \nu_V := \phi\psi\mu$, where ϕ (resp. ψ) is the right (resp. left) eigenfunction of P^V associated with $e^{\Lambda(V)}$ given in Lemma 3.1(a) *verifying $\mu(\phi\psi) = 1$.*

Proof. Recall at first that

$$J(\nu) = \inf\left\{J^{(2)}(Q);\ Q \in M_1^s(E^2), Q(A \times E) = \nu(A),\ \forall A \in \mathscr{B}(E)\right\}, \tag{3.7}$$

where $J^{(2)}(Q)$ is given in (2.12) (cf. [7,5]).

"$\Longleftarrow$" Let $\mathbb{Q}^V(\mathrm{d}x, \mathrm{d}y) = v_V(\mathrm{d}x)Q^V(x, \mathrm{d}y)$. By definition (2.12), we have

$$J^{(2)}(\mathbb{Q}^V) = \mathbb{E}^{\mathbb{Q}^V} \log \frac{Q^V(x, \mathrm{d}y)}{P(x, \mathrm{d}y)} = \mathbb{E}^{\mathbb{Q}^V} \log \frac{\phi(y)}{\mathrm{e}^{\Lambda(V)}\phi(x)} \cdot \mathrm{e}^{V(x)}$$
$$= \int \log \frac{\mathrm{e}^{V(x)}}{\mathrm{e}^{\Lambda(V)}} \, \mathrm{d}v_V(x) = \langle V, v_V \rangle - \Lambda(V). \tag{3.8}$$

By (3.7), $J(v_V) \leqslant \langle V, v_V \rangle - \Lambda(V)$ and the equality holds by (3.1).

"$\Longrightarrow$" It is well known from the convex analysis that

$$J(v) = \langle v, V \rangle - \Lambda(V) \Longleftrightarrow v \in \partial\Lambda(V), \tag{3.9}$$

where $\partial\Lambda(V)$ denotes the set of sub-differentials of $\Lambda(\cdot)$ at V (which is contained in the topological dual space $(b\mathscr{B}(E))'$ to which $M_1(E)$ is embedded). Since $v_V \in \partial\Lambda(V)$ (by the sufficiency above) and $\Lambda(V)$ is Gateaux-differentiable on $b\mathscr{B}$ by Lemma 3.3, $\partial\Lambda(V)$ is the singleton $\{v_V\}$. $\square$

The following lemma is a main result in [15], which will be crucial in the proof of the lower bound in Theorem 2.2.

Lemma 3.5 (*Lei and Wu [15, Theorem 2.1]*). *Given a stationary sequence $(X_i)_{i \in \mathbb{N}}$ valued in E such that $\mu(\mathrm{d}x) = \mathbb{P}(X_i \in \mathrm{d}x) = f(x)\,\mathrm{d}x$. Let $(\phi_k)_{k \geqslant 1}$ be the ϕ-mixing coefficient of $(X_i)_{i \in \mathbb{N}}$. Assume* (1.3) *and*

$$S_\phi := \sum_{k=1}^{\infty} \phi_k < +\infty. \tag{3.10}$$

Let D_n^* be given by (1.2). *Then $D_n^* \to 0$ exponentially as $n \to \infty$, i.e.,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(D_n^* > \delta) < 0, \quad \forall \delta > 0.$$

Corollary 3.6. *If P is a Doeblin recurrent* [17] *Markov kernel on E with the unique invariant probability measure* $\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$, *then*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}_\mu(D_n^* > \delta) < 0, \quad \forall \delta > 0.$$

Proof. If P is moreover aperiodic, then $S_\phi < +\infty$ (well known, see the proof of Theorem 2.3(c) in Section 6) and this corollary follows directly from Lemma 3.5. Now assume that P is of period $d > 1$. By the classical theory of Markov chains in [17], we have the following cyclic decomposition: $E = \mathscr{N} \cup E_1 \cup \cdots \cup E_d$ where $\mu(\mathscr{N}) = 0$ and

(i) $\mathscr{N}, E_1, \ldots, E_d$ are disjoint;
(ii) $P(x, E_{i+1}) = 1, \ \forall x \in E_i$ (here $E_{d+1} := E_1$);
(iii) there are $C > 0$ and $r \in (0, 1)$ such that

$$\sup_{x \in E_i} \|P^{nd}(x, \cdot) - \mu_i\|_{\mathrm{TV}} \leqslant Cr^n, \quad \forall n \geqslant 0, \ i = 1, \ldots, d,$$

where $f_i = f1_{E_i}d$ and $\mu_i = d1_{E_i}\mu = f_i\,\mathrm{d}x$. Let

$$f^*_{n,d}(x) := \frac{1}{n}\sum_{k=0}^{n-1} K_{h_n}(x - X_{dk}).$$

Since $P^d|_{E_i}$ is Doeblin recurrent and aperiodic on E_i by property (iii) above, we have by Lemma 3.5,

$$\limsup_{n\to\infty}\frac{1}{n}\log \mathbb{P}_{\mu_i}(\|f^*_{n,d}\circ\theta^j - f_{i+j}\|_1 > \delta) < 0, \quad \forall \delta > 0$$

for all $i,j = 1,\ldots,d$ where $i+j := i+j(\mathrm{mod}\,d)$. As $f^*_{nd}(x) = \frac{1}{d}\sum_{j=1}^d f^*_{n,d}\circ\theta^j$ and $f = \frac{1}{d}\sum_{j=1}^d f_{i+j}$, then we get for any $\delta > 0$ and $i = 1,\ldots,d$

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{nd}\log \mathbb{P}_{\mu_i}(\|f^*_{nd} - f\|_1 > \delta)\\ &\quad\leqslant \limsup_{n\to\infty}\frac{1}{nd}\log\sum_{j=1}^d \mathbb{P}_{\mu_i}(\|f^*_{n,d}\circ\theta^j - f_{i+j}\|_1 > \delta) < 0,\end{aligned}$$

where the desired result follows. □

Lemma 3.7. *Under* H1, *we have*:

(a) *for any* $k \geqslant 1$, *there exists some* $\delta > 0$ *such that*

$$\sup_{|t|\leqslant\delta}\ \sup_{\|V\|\leqslant 1}\left|\frac{\mathrm{d}^k}{\mathrm{d}t^k}\Lambda(tV)\right| < +\infty$$

and for every $V \in b\mathscr{B}(E)$, $\Lambda''(tV)|_{t=0} = \sigma^2(V)$;

(b) *the rate function* J_V *given in* (2.7) *satisfies*

$$J_V(r) = \begin{cases}\sup_{t\in\mathbb{R}}(t[(r+\mu(V)] - \Lambda(tV)), & \forall r \in \mathbb{R},\\ \sup_{t\geqslant 0}(t[(r+\mu(V)] - \Lambda(tV)), & \forall r \geqslant 0\end{cases} \tag{3.11}$$

and J_V *is strictly convex on* $[J_V < +\infty]^0 = (a,b)$ *where* $a = \lim_{t\to-\infty}\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(tV) - \mu(V)$ *and* $b = \lim_{t\to+\infty}\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(tV) - \mu(V)$ (*in particular* J_V *is strictly increasing and continuous in* $[0,b)$); *moreover*

$$\lim_{r\to 0+}\frac{J_V(r)}{r^2} = \frac{1}{2\sigma^2(V)} \in (0,+\infty].$$

Proof. (a) We shall follow the approach in [21], in which it is assumed that 1 is the unique isolated eigenvalue $z \in \mathbb{C}$ of P in $b\mathscr{B}(E)$ such that $|z| = 1$. Under H1, the last assumption is satisfied if P is aperiodic. Let us see how to bypass this assumption.

Under H1, recall the cyclic decomposition $E = \mathcal{N}\cup\bigcup_{i=1}^d E_i$ in the proof of Corollary 3.6 above. Let us consider $P^d|_{E_i}$ which is Doeblin recurrent on E_i, aperiodic, with the unique invariant probability measure μ_i. Hence 1 is the unique isolated eigenvalue $z \in \mathbb{C}$ of $P^d|_{E_i}$ in $b\mathscr{B}(E_i)$ such that $|z| = 1$.

For each $V \in b\mathscr{B}(E)$, consider the following operator acting on $b\mathscr{B}(E_1)$:

$$R^V f(x) := \mathbb{E}^x f(X_d) \mathrm{e}^{\sum_{k=0}^{d-1} V(X_k)} = (P^V)^d|_{E_1} f(x), \quad \forall x \in E_1.$$

It is obvious that the spectral radius $r_{\mathrm{sp}}(R^V)$ of R^V in $b\mathscr{B}(E_1)$ is not greater than $r_{\mathrm{sp}}((P^V)^d) = \mathrm{e}^{\mathrm{d}\Lambda(V)}$. On the other hand, by the LDP in [5] for any initial measure and the fact that $(P^V)^d 1_{E_1^c} = 0$ on E_1, we have

$$\begin{aligned} \log r_{\mathrm{sp}}(R^V) &\geqslant \lim_{n\to\infty} \frac{1}{n} \log \mu_1[(P^V)^{nd} 1_{E_1}] \\ &= \lim_{n\to\infty} \frac{1}{n} \log \mu_1[(P^V)^{nd} 1] \\ &= \mathrm{d}\Lambda(V). \end{aligned}$$

Thus $r_{\mathrm{sp}}(R^V) = \mathrm{e}^{\mathrm{d}\Lambda(V)}$.

As in [21], we will apply the analytical perturbation theory of Kato [12]. For each $z \in \mathbb{C}$, consider R^{zV} acting on the complexified space $b_{\mathbb{C}}\mathscr{B}(E_i)$, which is analytical in z in the sense of [12]. Then for any $\eta \in (0, 1/2)$ sufficiently small, there exists $\delta > 0$ and $C > 0$ such that for all $V \in b\mathscr{B}(E)$ with $\|V\| \leqslant 1$,

(1) the eigenvalue $\lambda_{\max}(R^{zV})$ of R^{zV} with the largest modulus is isolated in the spectrum of R^{zV} and $|\lambda_{\max}(R^{zV}) - 1| \leqslant \eta$ for $|z| \leqslant 2\delta$;
(2) for all $|z| \leqslant 2\delta$, the eigenprojection $E(z, V)$ of R^{zV} associated with $\lambda_{\max}(R^{zV})$ is unidimensional and

$$\|E(z, V)1_{E_1} - 1_{E_1}\| < 1/2, \ \ \|(R^{zV})^n (I - E(z, V))\| \leqslant C(1 - 2\eta)^{nd}, \quad \forall n;$$

(3) $z \to \lambda_{\max}(R^{zV})$ and $z \to E(z, V)f$ is analytic in z for $|z| \leqslant 2\delta$;

where properties (1) and (2) follow by [12, Chapter IV, Theorem 3.16] and property (3) by [12, Chapter VII, Theorem 1.8].

Then $\Lambda(zV) := \frac{1}{d} \log \lambda_{\max}(R^{zV})$ is analytic for $|z| \leqslant 2\delta$ and coincides with $\Lambda(tV)$ when $z = t \in [-2\delta, 2\delta] \subset \mathbb{R}$.

Let $\Lambda_n(zV) := \frac{1}{nd} \log \mathbb{E}^{\mu_1} \exp(\sum_{k=0}^{nd-1} zV(X_k)) = \frac{1}{nd} \log\langle 1, (R^{zV})^n 1\rangle_{\mu_1}$. By the properties (1) and (2) above, we have

$$\langle 1, (R^{zV})^n 1\rangle_{\mu_1} = \mathrm{e}^{nd\Lambda(zV)} \langle 1, E(z, V)1\rangle_{\mu_1} + \mathrm{O}((1 - 2\eta)^{nd}),$$

where it follows that $\Lambda_n(zV) \to \Lambda(zV)$ uniformly over $z : |z| \leqslant 2\delta$ and $V : \|V\| \leqslant 1$. Thus by Cauchy's theorem and property (3) above,

$$\sup_{\|V\|\leqslant 1} \sup_{|z|\leqslant\delta} \left| \frac{\mathrm{d}^k}{\mathrm{d}z^k} \Lambda(zV) \right| < +\infty,$$

$$\sup_{\|V\|\leqslant 1} \sup_{|z|\leqslant\delta} \left| \frac{\mathrm{d}^k}{\mathrm{d}z^k} \Lambda_n(zV) - \frac{\mathrm{d}^k}{\mathrm{d}z^k} \Lambda(zV) \right| \to 0.$$

Applying the above estimate to $k=2$ and notice that $\mathbb{E}^{\mu_i}\sum_{k=1}^{d}V(X_k)=\mathrm{d}\mu(V)$,

$$\begin{aligned}\Lambda_n''(tV)|_{t=0} &= \frac{1}{nd}\,\mathbb{E}^{\mu_1}\left(\sum_{k=0}^{nd-1}V(X_k)-\mathbb{E}^{\mu_1}\sum_{k=0}^{nd-1}V(X_k)\right)^2\\ &\to \mathrm{Var}_{\mathbb{P}_{\mu_1}}\left(\sum_{k=0}^{d-1}V(X_k)\right)+2\sum_{n=1}^{\infty}\mathrm{Cov}_{\mathbb{P}_{\mu_1}}\left(\sum_{k=0}^{d-1}V(X_k),\sum_{k=0}^{d-1}V(X_{nd+k})\right).\end{aligned}$$

From the cyclic decomposition, we see that the last quantity above is exactly $\sigma^2(V)$. Thus $\Lambda''(tV)|_{t=0}=\sigma^2(V)$.

(b) By the LDP of L_n in [5] and the Laplace principle due to Varadhan, we have for all $t\in\mathbb{R}$,

$$\Lambda(t[V-\mu(V)])=\sup\{\nu(tV)-t\mu(V)-J(\nu);\ \nu\in M_1(E)\}=\sup_{r\in\mathbb{R}}\{tr-J_V(r)\},$$

Hence the Legendre–Fenchel theorem gives us

$$J_V(r)=\sup_{t\in\mathbb{R}}\{tr-\Lambda(t[V-\mu(V)])\}=\sup_{t\in\mathbb{R}}\{t(r+\mu(V)]-\Lambda(tV)\},\quad \forall r\in\mathbb{R}$$

for $\Lambda(t[V-\mu(V)])=\Lambda(tV)-t\mu(V)$. When $r\geqslant 0$, since $\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(tV)|_{t=0}=\mu(V)$, the supremum above can be taken only for $t\geqslant 0$. Then (3.11) is proved.

All other properties of $J_V(r)=\sup_{t\in\mathbb{R}}(tr-\Lambda(t[V-\mu(V)]))$ are easy consequences of the elementary convex analysis. $\square$

Lemma 3.8 (*Bishop–Phelps, cf. [20] or [22]*). *Assume Λ is a convex real function on a Banach space Y. Assume $x_o\in Y'$ (the topological dual space) satisfies*:

$$\exists c\in\mathbb{R}:\Lambda(y)\geqslant\langle x_0,y\rangle-c,\quad\forall y\in Y$$

then $\forall y\in Y,\forall\varepsilon>0,\exists y'\in Y,x'\in\partial\Lambda(y')$, *such that*

$$\|x'-x_0\|\leqslant\varepsilon,\|y'-y\|\leqslant\frac{1}{\varepsilon}(\Lambda(y)-\langle x_0,y\rangle+\Lambda^*(x_0)),$$

where $\Lambda^*(x):=\sup\{\langle x,y\rangle-\Lambda(y)\mid y\in Y\})$, $\forall x\in Y'$, *is the Legendre transformation of* $\Lambda(y)$.

4. Proof of Theorem 2.1

The desired LDP of f_n^* in $(L^1(\mathbb{R}^d),\sigma(L^1,L^\infty))$ is equivalent to the LDP of $f_n^*(x)\,\mathrm{d}x$ on $M_1(\mathbb{R}^d)$ w.r.t. the τ-topology $\sigma(M_1(\mathbb{R}^d),b\mathscr{B})$. Since $\Lambda(V1_E)$ is Gateaux-differentiable on $b\mathscr{B}$ by Lemma 3.3(a), by the abstract Gärtner–Ellis theorem [22, p. 290, Theorem 2.7], it is enough to show that for each $V\in b\mathscr{B}$,

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\log\inf_{x\in E}\mathbb{E}^x\exp\left(n\int_{\mathbb{R}^d}f_n^*(y)V(y)\,\mathrm{d}y\right)\\ &\quad=\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{E}^x\exp\left(n\int_{\mathbb{R}^d}f_n^*(y)V(y)\,\mathrm{d}y\right)=\Lambda(V1_E)\end{aligned}\tag{4.1}$$

and $\Lambda(V1_E)$ is monotonely continuous at 0, i.e., if (V_n) is a sequence in $b\mathscr{B}$ decreasing pointwise to 0 over $\mathbb{R}^d$, then $\Lambda(V_n 1_E) \to 0$.

The last condition is satisfied by Lemma 3.3(b). It remains to verify (4.1). Put $V_n = (K_{h_n} * V)1_E$, then $\|V_n\| \leqslant \|V\|$ and,

$$n\int_{\mathbb{R}^d} f_n^*(y)V(y)\,\mathrm{d}y = \sum_{k=0}^{n-1} V_n(X_k).$$

Consequently letting ϕ_n be the right eigenfunction of P^{V_n} associated with $\mathrm{e}^{\Lambda(V_n)}$, and $C := M\mathrm{e}^{3N\|V\|}$, we have by Lemma 3.1(a) that for each $x \in E$,

$$\begin{aligned}\mathbb{E}^x \exp\left(n\int_{\mathbb{R}^d} f_n^*(y)V(y)\,\mathrm{d}y\right) &\leqslant C\mathbb{E}^x\, \frac{\phi_n(X_n)}{\phi_n(x)} \exp\left(\sum_{k=0}^{n-1} V_n(X_k)\right)\\ &= C\,\frac{(P^{V_n})^n\phi_n(x)}{\phi_n(x)} = C\mathrm{e}^{n\Lambda(V_n)}\end{aligned}$$

and similarly

$$\mathbb{E}^x \exp\left(n\int_{\mathbb{R}^d} f_n^*(y)V(y)\,\mathrm{d}y\right) \geqslant \frac{1}{C}\,\mathbb{E}^x\, \frac{\phi_n(X_n)}{\phi_n(x)} \exp\left(\sum_{k=0}^{n-1} V_n(X_k)\right) = \frac{1}{C}\,\mathrm{e}^{n\Lambda(V_n)}.$$

Noting that $V_n \to V1_E$, $\mathrm{d}x$-*a.e.*, we have $\Lambda(V_n) \to \Lambda(V1_E)$ by Lemma 3.3(b). Thus the two estimations above yield the desired relation (4.1).

5. Proof of Theorem 2.2

Part 1 (*Large deviation upper bound*). This is an easy consequence of Theorem 2.1. In fact, for any $g \in L^1(\mathbb{R}^d)$ and δ fixed, as $\{\tilde{g} \in L^1(\mathbb{R}^d);\ \|\tilde{g} - g\|_1 \leqslant \delta\}$ is closed in the weak topology $\sigma(L^1, L^\infty)$, then by Theorem 2.1,

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathbb{R}^d} \mathbb{P}_x(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} \leqslant \delta) \leqslant -\inf_{\tilde{g}:\|\tilde{g}-g\|_1\leqslant\delta} J(\tilde{g}).$$

Letting $\delta \to 0$, we get the desired result by the lower semi-continuity of J (which follows from (3.1)).

Part 2 (*Large deviation lower bound*). It is enough to prove that for any $g \in \mathscr{P}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \geqslant -J(g), \quad \forall \delta > 0.$$

Its proof, more difficult, is divided into three steps.

Step 1: We claim that it is enough to show that for any $g \in \mathscr{P}$ and $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \geqslant -J(g), \quad \mu\text{-}a.s.\ x \in A \tag{5.1}$$

for some $A \in \mathscr{B}(E)$ charged by μ. Indeed, if (5.1) is true, then by Egorov's lemma, there is some measurable $U \subset A$ with $\mu(U) > 0$ such that

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in U} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \geqslant -J(g).$$

Let $\tau_U := \inf\{n \geqslant 1;\ X_n \in U\}$ be the first hitting time to U. By H1, we have $\frac{1}{N}\sum_{k=1}^{N} P^k(x,\cdot) \geqslant \frac{1}{M}\mu(\cdot)$, then

$$\inf_{x\in E} \mathbb{P}_x(\tau_U \leqslant N) \geqslant \inf_{x\in E} \mathbb{E}^x \frac{\sum_{k=1}^{N} 1_U(X_k)}{N} \geqslant \frac{\mu(U)}{M} > 0.$$

Since

$$f_n^* \circ \theta^k := \frac{1}{n} \sum_{i=k}^{k+n-1} \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right)$$

we have $\|f_n^* - f_n^* \circ \theta_{\tau_U}\|_1 \leqslant \frac{2N}{n}$ on $[\tau_U \leqslant N]$. Thus by the strong Markov property, we have for $n \geqslant N$ such that $2N/n < \delta/2$,

$$\inf_{x\in E} \mathbb{P}_x(\|f_n^* - g\|_1 < \delta) \geqslant \inf_{x\in E} \mathbb{P}_x(\tau_U \leqslant N) \cdot \inf_{y\in U} \mathbb{P}_y\left(\|f_n^* - g\|_1 < \frac{\delta}{2}\right),$$

where the desired uniform lower bound follows from (5.1).

Step 2: For "$g\,\mathrm{d}x = v_V$" case. The idea of this step is to use change of measure. Given $V \in b\mathscr{B}$, let Q^V be the transition kernel defined in Lemma 3.1 and $v_V = \phi\psi\mu$. From Lemma 3.1, we know that Q^V is Doeblin recurrent.

Let $\mathbb{Q}^V_{\omega(0)}$ be the law of the Markov process with transition kernel Q^V and the initial point $\omega(0)$, which is v_V-*a.s.* well-defined on $\Omega = E^{\mathbb{N}}$, and $\mathbb{Q}^V := \int \mathbb{Q}^V_{\omega(0)}\,\mathrm{d}v_V(\omega(0))$. Denoting by $\xi(\omega)$ the density of $\mathbb{Q}^V_{\omega(0)}$ w.r.t. $\mathbb{P}_{\omega(0)}$ on $\sigma(X_1)$, we have for μ-*a.s.* $\omega(0)$,

$$\left.\frac{\mathrm{d}\mathbb{Q}^V_{\omega(0)}(\mathrm{d}\omega_1,\ldots,\mathrm{d}\omega_n)}{\mathrm{d}\mathbb{P}_{\omega(0)}}\right|_{\mathscr{F}_n} = \exp\left(\sum_{k=0}^{n-1} \log\, \xi(\theta^k\omega)\right)$$

and $\mathbb{E}^{\mathbb{Q}^V} \log\, \xi = J^{(2)}(\mathbb{Q}^V|_{\mathscr{F}_1}) = J(v_V)$ by Lemma 3.4. For any $\varepsilon > 0$, putting

$$W_n := \{\omega : \|f_n^*(\omega) - g\|_1 < \delta\}, \quad D_{n,\varepsilon} := \left\{\omega : \frac{1}{n}\sum_{n=0}^{n-1} \log\, \xi(\theta^k\omega) \leqslant J(v_V) + \varepsilon\right\},$$

we have for μ-*a.s.* $\omega(0)$,

$$\begin{aligned} \mathbb{P}_{\omega(0)}(W_n) &\geqslant \int_{W_n} \exp\left(-\sum_{k=0}^{n-1} \log\, \xi(\theta^k\omega)\right) \mathrm{d}\mathbb{Q}^V_{\omega(0)} \\ &\geqslant \exp[-n(J(v_V) + \varepsilon)] \cdot \mathbb{Q}^V_{\omega(0)}\big(W_n \cap D_{n,\varepsilon}\big). \end{aligned} \tag{5.2}$$

So to get (5.1), it remains to show that $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$ and $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$ for μ-*a.s.* $\omega(0)$, as n goes to infinity (for any $\varepsilon > 0$).

By the ergodic theorem and the Fubini theorem, we have for $v_V \sim \mu$-*a.s.* $\omega(0)$,

$$\frac{1}{n}\sum_{k=0}^{n-1} \log\, \xi(\theta_k\omega) \to \mathbb{E}^{\mathbb{Q}^V} \log\, \xi = J(v_V), \quad \mathbb{Q}^V_{\omega(0)}\text{-}a.s.,$$

where follows $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$. For the second limit, applying the crucial Corollary 3.6 to $((X_n), \mathbb{Q}^V)$ (where the condition is satisfied because $((X_n), \mathbb{Q}^V)$ is Doeblin recurrent by Lemma 3.1), we have

$$\mathbb{Q}^V(W_n^c) \to 0 \quad \text{exponentially rapidly.}$$

Then by the Borel–Cantelli lemma,

$$\mathbb{Q}^V(W_n^c, \text{ infinitely often}) = 0.$$

By Fubini's theorem, $\mathbb{Q}^V_{\omega(0)}(W_n^c, \text{ infinitely often}) = 0$, for $v_V \sim \mu$-*a.s.* $\omega(0)$.

Step 3: The general case. By Steps 1 and 2, it remains to show:

Claim. $\forall v = g\,\mathrm{d}x \in M_1(\mathbb{R}^d)$ *satisfies* $J(g) < +\infty$, *there exists a sequence of* $(v_n) := (v_{V_n})$, *such that* $\|v_{V_n} - v\|_{\mathrm{TV}} \to 0$ *and* $\limsup_{n\to\infty} J(v_{V_n}) \leqslant J(v)$.

Let us construct this sequence by means of Bishop–Phelps theorem (Lemma 3.8). For any $n \geqslant 1$, we choose $\tilde{V}_n \in b\mathscr{B}(E)$ such that $J(v) < \langle v, \tilde{V}_n\rangle - \Lambda(\tilde{V}_n) + \frac{1}{n}$ (by (3.1)). By Lemma 3.8, for each $\tilde{V}_n$ and $\varepsilon_n = \frac{1}{n(\|\tilde{V}_n\|+1)}$, we can find $V_n \in b\mathscr{B}(E)$, $v_{V_n} \in \partial\Lambda(V_n)$ (which is a singleton $\{v_{V_n}\}$ by the proof of Lemma 3.4), such that

$$\|v_{V_n} - v\|_{\mathrm{TV}} \leqslant \varepsilon_n, \quad \|\tilde{V}_n - V_n\| \leqslant \frac{1}{\varepsilon_n}(\Lambda(\tilde{V}_n) - \langle v, \tilde{V}_n\rangle + J(v)).$$

So

$$\langle v_{V_n} - v, V_n\rangle \leqslant \|v_{V_n} - v\|_{\mathrm{TV}} \cdot \|V_n - \tilde{V}_n\| + \|v_{V_n} - v\|_{\mathrm{TV}} \cdot \|\tilde{V}_n\| \leqslant \frac{2}{n}.$$

As $\partial\Lambda(V_n) = \{v_{V_n}\}$, we have,

$$J(v_{V_n}) = \langle v_{V_n}, V_n\rangle - \Lambda(V_n) = \langle v_{V_n} - v, V_n\rangle + \langle v, V_n\rangle - \Lambda(V_n) \leqslant \frac{2}{n} + J(v).$$

This proves the claim. The proof of the theorem is completed.

6. Proof of Theorem 2.3

6.1. *Proof of part (a) in Theorem 2.3*

Its proof is divided into two parts.

Part 1 (*Lower bound in* (2.3)). The lower bound is an easy consequence of Theorem 2.1. Actually, as $\{g \in L^1(\mathbb{R}^d);\ \|g - f\|_1 > \delta\}$ is open in the weak topology $\sigma(L^1, L^\infty)$, we have by Theorem 2.1,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in E} \mathbb{P}_x(\|f_n^* - f\|_1 > \delta) \geqslant - \inf_{g:\ \|g-f\|_1 > \delta} J(g) = -I(\delta).$$

Part 2 (*Upper bound in* (2.3)). The proof of the upper bound is much more difficult, and it is divided into three steps, where the first two steps are similar to [6] and the third one is inspired by [16].

Step 1 (*Approximation of* K). The purpose of this step is to show that we can reduce to the case where $K = \frac{1}{|A|} 1_A$, $A := \prod_{i=1}^d [x_i, x_i + a_i)$ is a rectangle (here $|A|$ denotes the Lebesgue measure of $A \in \mathscr{B}$).

Given $\varepsilon > 0$, we can find finite positive constants q, m, $b_1, \ldots, b_m$ and disjoint finite rectangles $A_1, \ldots, A_m$ in $\mathbb{R}^d$ of form $\prod_{i=1}^d [x_i, x_i + a_i)$ such that the function

$$K^{(\varepsilon)}(x) = \sum_{j=1}^m b_j I_{A_j}(x)$$

satisfies $\int K^{(\varepsilon)}(x)\,\mathrm{d}x = 1$, $K^{(\varepsilon)} \leqslant q$ and $\int |K(x) - K^{(\varepsilon)}|\,\mathrm{d}x < \varepsilon$. Define

$$f_n^{(\varepsilon),*} := K_{h_n}^{(\varepsilon)} * \mathrm{d}L_n = \frac{1}{n} \sum_{k=0}^{n-1} \frac{1}{h_n^d} K^{(\varepsilon)}\left(\frac{x - X_k}{h_n}\right).$$

Then

$$\begin{aligned} \int |f_n^*(x) - f_n^{(\varepsilon),*}(x)|\,\mathrm{d}x &\leqslant \int h_n^{-d} \int \left| K^{(\varepsilon)}\left(\frac{x-y}{h_n}\right) - K\left(\frac{x-y}{h_n}\right)\right| L_n(\mathrm{d}y)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} |K^{\varepsilon} - K|(z)\,\mathrm{d}z \leqslant \varepsilon. \end{aligned}$$

Thus by the approximation lemma in large deviations [4] (more precisely, by the same cycle of idea), it is enough to prove that $f_n^{(\varepsilon),*}$ satisfies the upper bound in (2.3).

Let $K^j = \frac{1}{|A_j|} 1_{A_j}$, then $K^{(\varepsilon)} = \sum_{j=1}^m \lambda_j K^j$ where $\sum_{j=1}^m \lambda_j = 1$ and $\lambda_j > 0$. Consequently,

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^{(\varepsilon),*} - f\|_1 > \delta) \\ &\quad \leqslant \limsup_{n\to\infty} \frac{1}{n} \log \sum_{j=1}^m \sup_{x\in E} \mathbb{P}_x(\|K_{h_n}^j * \mathrm{d}L_n - f\|_1 > \delta) \\ &\quad = \max_{1\leqslant j\leqslant m} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|K_{h_n}^j * \mathrm{d}L_n - f\|_1 > \delta). \end{aligned}$$

Thus for the upper bound in (2.3), we may (and will) assume that $K = \frac{1}{|A|} 1_A$ where $A := \prod_{i=1}^d [x_i, x_i + a_i)$.

Step 2 (*Method of partition*). Fix such a rectangle $A := \prod_{i=1}^d [x_i, x_i + a_i)$ and $K = \frac{1}{|A|} 1_A$, and let $0 < \varepsilon < \delta/4$ be arbitrary. Since $K_{h_n} * f \to f$ in $L^1(\mathbb{R}^d)$, then it is enough to show that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in E} \mathbb{P}_x(\|f_n^* - K_{h_n} * f\|_1 > \delta) \leqslant -I(\delta-). \tag{6.1}$$

Note that

$$\int |f_n^*(x) - K_{h_n} * f(x)|\,\mathrm{d}x \leqslant \int \left| \frac{1}{|A|h_n^d} \int_{x+h_nA} L_n(\mathrm{d}y) - \frac{1}{|A|h_n^d} \int_{x+h_nA} f(y)\,\mathrm{d}y \right| \mathrm{d}x$$
$$\leqslant \frac{1}{|A|h_n^d} \int |L_n(x+h_nA) - \mu(x+h_nA)|\,\mathrm{d}x.$$

Consider the partition of $\mathbb{R}^d$ into sets B that are d-fold products of intervals of the form $[\frac{(i-1)h_n}{p}, \frac{ih_n}{p})$, where $i \in \mathbb{Z}$, and $p \in \mathbb{N}^*$ such that $\min_i a_i \geqslant \frac{2}{p}$. Call the partition Ψ.

Let $A^* = \prod_{i=1}^d [x_i + \frac{1}{p}, x_i + a_i - \frac{1}{p})$. We have

$$C_x := (x+h_nA) \backslash \bigcup_{B\in\Psi, B\subseteq x+h_nA} B \subseteq x + h_n(A\backslash A^*).$$

Consequently,

$$\int |f_n^*(x) - K_{h_n} * f(x)|\,\mathrm{d}x$$
$$\leqslant \frac{1}{|A|h_n^d} \int \sum_{B\in\Psi, B\subseteq x+hA} |L_n(B) - \mu(B)|\,\mathrm{d}x + \frac{1}{|A|h_n^d} \int \{\mu(C_x) + L_n(C_x)\}\,\mathrm{d}x. \quad (6.2)$$

Using the fact that for any set $C \in \mathscr{B}$, $h > 0$ and any probability measure v on $\mathbb{R}^d$,

$$\int v(x+hC)\,\mathrm{d}x = |hC| = h^d|C|$$

(by Fubini), the last term in (6.2) is bounded from above by

$$\frac{1}{|A|h_n^d} 2h_n^d |A\backslash A^*| = \frac{2}{|A|}\left(\prod_{i=1}^d a_i - \prod_{i=1}^d \left(a_i - \frac{2}{p}\right)\right)$$
$$= 2\left(1 - \prod_{i=1}^d \left(1 - \frac{2}{pa_i}\right)\right) \leqslant \varepsilon$$

once if p verifies

$$\min_i a_i \geqslant \frac{2}{p}, \quad 2\left(1 - \prod_{i=1}^d \left(1 - \frac{2}{pa_i}\right)\right) \leqslant \varepsilon.$$

We fix such p which is independent of n.

For any finite constant $R > 0$, letting $S_{OR} := \{x \in \mathbb{R}^d;\ |x| \leqslant R\}$, we can bound the first term at the r.h.s. of (6.2) from above by

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\phi} |L_n(B) - \mu(B)| \frac{1}{|A|h_n^d} \int_{B\subseteq x+h_nA} \mathrm{d}x$$
$$+ \frac{1}{|A|h_n^d} \int_{B\subseteq x+h_nA} \mathrm{d}x \{L_n(S_{OR}^c) - \mu(S_{OR}^c) + 2\mu(S_{OR}^c)\}.$$

Clearly, $h_n^{-d} \int_{B\subseteq x+h_nA} \mathrm{d}x \leqslant |A|$, and $\mu(S_{OR}^c) < \varepsilon/2$ for all $R \geqslant R_0$.

By Lemma 3.2, we have for all $t>0$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\{L_n(S_{OR}^c)-\mu(S_{OR}^c)>\varepsilon\}$$
$$\leqslant -J_{S_{OR}^c}(\varepsilon)\leqslant -(t[\varepsilon+\mu(S_{OR}^c)]-\Lambda(t1_{S_{OR}^c})).$$

Since $\lim_{R\to\infty}\Lambda(t1_{S_{OR}^c})=0$ by Lemma 3.3, then for any $L>0$, the l.h.s. above is bounded from above by $-L$ for all R large enough, say $R\geqslant R_1$. Fix $R\geqslant R_0\vee R_1$ below. Summarizing those estimations we obtain

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\int |f_n^*(x)-K_{h_n}*f(x)|\,\mathrm{d}x>\delta\right)$$
$$\leqslant(-L)\vee\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\sum_{B\in\Psi,B\cap S_{OR}\neq\emptyset}|L_n(B)-\mu(B)|>\delta-3\varepsilon\right). \quad (6.3)$$

Step 3: It remains to control the last term in (6.3). Set

$$\tilde{\Psi}=\{B;B\in\Psi,B\cap S_{OR}\neq\emptyset\}\cup\{C\},\quad C:=\left(\bigcup_{B\in\tilde{\Psi}}B\right)^c$$

and $\mathscr{B}(\tilde{\Psi})=\sigma\{B;B\in\tilde{\Psi}\}$, the σ-field generated by $\tilde{\Psi}$. Regarding L_n and μ as probability measures on $\mathscr{B}(\tilde{\Psi})$, and denoting the total variation of $L_n-\mu$ on $\mathscr{B}(\tilde{\Psi})$ by $\|L_n-\mu\|_{\mathscr{B}(\tilde{\Psi})}$, we have

$$\sum_{B\in\Psi,B\cap S_{OR}\neq\emptyset}|L_n(B)-\mu(B)|\leqslant\|L_n-\mu\|_{\mathscr{B}(\tilde{\Psi})}=\max_{V\in\{-1,1\}^{\tilde{\Psi}}}(L_n(V)-\mu(V)),$$

where $\{-1,1\}^{\tilde{\Psi}}$ denotes the set of all $\mathscr{B}(\tilde{\Psi})$-measurable functions with values in $\{-1,1\}$ (which can be identified as the set of functions from $\tilde{\Psi}$ to $\{-1,1\}$). Therefore, for any $r>0$ fixed,

$$\mathbb{P}_x\left(\sum_{B\in\Psi,B\cap S_OR\neq\emptyset}|L_n(B)-\mu(B)|>r\right)\leqslant\mathbb{P}_x\left(\max_{V\in\{-1,1\}^{\mathscr{B}(\tilde{\Psi})}}L_n(V)-\mu(V)>r\right)$$
$$\leqslant\sum_{V\in\{-1,1\}^{\mathscr{B}(\tilde{\Psi})}}\mathbb{P}_x(L_n(V)-\mu(V)>r).$$

At first by Lemma 3.2, for each $V\in\{-1,1\}^{\tilde{\Psi}}$ and for all $0<\varepsilon<r$,

$$\sup_{x\in E}\mathbb{P}_x(L_n(V)-\mu(V)>r)\leqslant M\exp(-nJ_{V1_E}(r-\varepsilon)),\quad\forall n\geqslant\frac{4N}{\varepsilon}.$$

Secondly, the number of elements $\tilde{\Psi}$ is not greater than $(\frac{2Rp}{h_n}+2)^d+1=\mathrm{o}(n)$ by (1.3), then $\{-1,1\}^{\tilde{\Psi}}$ has $2^{\mathrm{o}(n)}$ elements for n large enough. Consequently letting $\mathbb{B}(1)$

be the unit ball in $L^{\infty}(\mu)$, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\sum_{B\in\Psi,B\cap S_O R\neq\emptyset}|L_n(B)-\mu(B)|>r\right)$$
$$\leqslant\limsup_{n\to\infty}\frac{1}{n}\log 2^{o(n)}M\sup_{V\in\mathbb{B}(1)}\exp(-nJ_V(r-\varepsilon))$$
$$=-\inf_{V\in\mathbb{B}(1)}J_V(r-\varepsilon),$$

where it follows by (6.3),

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in E}\mathbb{P}_x\left(\int|f_n^*(x)-K_{h_n}*f(x)|\,\mathrm{d}x>\delta\right)$$
$$\leqslant(-L)\vee\left(-\inf_{V\in\mathbb{B}(1)}J_V(\delta-4\varepsilon)\right).$$

As $L,\varepsilon>0$ are arbitrary and $\lim_{\varepsilon\to 0+}\inf_{V\in\mathbb{B}(1)}J_V(\delta-4\varepsilon)=I(\delta-)$ by (2.8), we obtain the desired (6.2) and then complete the proof of the upper bound in (2.3).

6.2. Proof of Part (b) in Theorem 2.3

Let $J(\nu/P)$ be the Donsker–Varadhan entropy of ν w.r.t. the Markov kernel P given by (1.5). We have for any $1\leqslant u\in b\mathscr{B}(E)$,

$$\int\log\frac{u}{P^N u}\,\mathrm{d}\nu=\sum_{k=0}^{N-1}\int\log\frac{P^k u}{PP^k u}\,\mathrm{d}\nu\leqslant NJ(\nu/P).$$

We get thus

$$NJ(\nu/P)\geqslant\sup_{1\leqslant u\in b\mathscr{B}(E)}\int\log\frac{u}{P^N u}\,\mathrm{d}\nu=J(\nu/P^N),\quad\forall\nu\in M_1(E),\ \forall N\geqslant 1.\tag{6.4}$$

By H1, $P^l(x,\cdot)\leqslant M\mu(\cdot)$. Then

$$J(\nu/P)\geqslant\frac{J(\nu/P^l)}{l}\geqslant\frac{1}{l}\left(\sup_{1\leqslant u\in b\mathscr{B}(E)}\int\log\frac{u}{\mu(u)}\,\mathrm{d}\nu-\log M\right)=\frac{h(\nu/\mu)-\log M}{l},$$

where $h(\nu/\mu)=\int\log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\mathrm{d}\nu$ if $\nu\ll\mu$ and $+\infty$ otherwise, is the relative entropy of ν w.r.t. μ (the last equality is the famous variational formula of relative entropy). Notice that in the i.i.d. case of common law μ, its transition is $P_0 f=\mu(f)$ and $h(\nu/\mu)=J(\nu/P_0)$. Hence

$$I^{iid}(\delta)=\inf\{h(\nu/\mu);\ \|\nu-\mu\|_{\mathrm{TV}}>\delta\},$$

where the desired inequality (2.5) follows.

6.3. Proof of Part (c) in Theorem 2.3

This follows from Rio's deviation inequality [19]. In fact using his inequality, we have (see [15] for details)

$$\mathbb{P}_\mu\left(\|f_n^* - f\|_1 - \mathbb{E}^\mu\|f_n^* - f\|_1 > \delta\right) \leqslant \exp\left(-\frac{n\delta^2}{8(1+2S_\phi)^2}\right), \quad \forall n \geqslant 1,\ \delta > 0,$$

where $S_\phi := \sum_{k=1}^{+\infty} \phi_k$ where ϕ_k is the ϕ-uniform mixing coefficient given in [19] or [15]. In the actual Markov context, we have

$$2\phi_k \leqslant \sup_{x,y\in E} \|P^k(x,\cdot) - P^k(y,\cdot)\|_{\mathrm{TV}}$$

and then $2S_\phi \leqslant S$, the quantity in (2.6). S is finite for aperiodic Doeblin recurrent Markov chain. Moreover by Lemma 3.5, $\mathbb{E}^\mu\|f_n^* - f\|_1 \to 0$. Thus by the lower bound in (2.3), Rio's estimate and the right continuity of $I(\delta)$, we get

$$-I(\delta) \leqslant -\frac{\delta^2}{8(1+S)^2},$$

where the desired inequality (2.6) follows.

7. Proof of Theorem 2.4

Lemma 7.1. *Given $V \in b\mathscr{B}(E)$. If T_n is an asymptotically consistent estimator of $\langle V,f\rangle := \int_E V(x)f(x)\,\mathrm{d}x$, i.e., for each $(P,\mu)\in\Theta$ (satisfying* H1 *and* $\mathrm{d}\mu(x) \ll \mathrm{d}x$), *$|\langle T_n, V\rangle - \langle f, V\rangle| \to 0$ in probability $\mathbb{P}_\mu$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\langle T_n - f, V\rangle > \delta) \geqslant -\inf\{J(g);\ \langle g - f, V\rangle > \delta\}. \tag{7.1}$$

Proof. It is enough to prove that the l.h.s. of (7.1) is $\geqslant -J(g)$ for every $g \in \mathscr{P}(E)$ such that $\langle g - f, V\rangle > \delta$ and $J(g) < +\infty$. By the Step 3 of the proof of Theorem 2.2, it suffices to prove it for $g\,\mathrm{d}x = v_{\tilde V}$ where $\tilde V \in b\mathscr{B}(E)$ is arbitrary. Its proof, completely parallel to the Step 2 in the proof of Theorem 2.2, is based on the fact that $(Q^{\tilde V}, v_{\tilde V}) \in \Theta$ again. It is so omitted. $\square$

Lemma 7.2. *Under* H1, *let $I(\cdot)$ be defined in* (2.4). *Then*

$$\lim_{r\to 0+} \frac{I(r)}{r^2} = \frac{1}{2\sup_{\|V\|\leqslant 1} \sigma^2(V)} = \frac{1}{8\sup_{A\in\mathscr{B}(E)} \sigma^2(1_A)}. \tag{7.2}$$

Proof. We shall only prove the first equality in (7.2) (the proof of the second is similar). By (2.8) and Lemma 3.7(b), for any $V \in b\mathscr{B}(E)$ with $\|V\| \leqslant 1$,

$$\limsup_{r\to 0} \frac{I(r)}{r^2} \leqslant \lim_{r\to 0} \frac{J_V(r+)}{r^2} = \frac{1}{2\sigma^2(V)},$$

where "$\leqslant$" in the first equality of (7.2) follows.

For the inverse inequality, let $L>1$ be arbitrary but fixed. For any $\delta>0$ small enough, we have by Lemma 3.7,

$$C(L\delta) := \sup_{t\in[0,L\delta]} \sup_{V\in\mathbb{B}(1)} \left|\frac{\mathrm{d}^3}{\mathrm{d}t^3}\Lambda(tV)\right| < +\infty.$$

Thus by the Taylor formula of order 3, we get for any $V \in \mathbb{B}(1)$ and $r \in (0,\delta]$,

$$\begin{aligned} J_V(r) &\geqslant \sup_{t\in[0,Lr]} (tr - \Lambda(t[V-\mu(V)])) \\ &\geqslant \sup_{t\in[0,Lr]} \left(tr - \frac{t^2\sigma^2(V)}{2}\right) - \frac{(Lr)^3}{6}\cdot C(L\delta) \\ &\geqslant r^2\left(L\wedge\sigma^{-2}(V) - \frac{[L\wedge\sigma^{-2}(V)]^2\sigma^2(V)}{2}\right) - \frac{(Lr)^3}{6}\cdot C(L\delta), \end{aligned}$$

where the last inequality is obtained by taking $t = r[L\wedge\sigma^{-2}(V)]$. Thus by (2.8),

$$\begin{aligned} \liminf_{r\to 0+} \frac{I(r)}{r^2} &= \liminf_{r\to 0+} \inf_{V\in\mathbb{B}(1)} \frac{J_V(r)}{r^2} \\ &\geqslant \min\left\{\inf_{V\in\mathbb{B}(1):\ \sigma^{-2}(V)\leqslant L} \frac{1}{2\sigma^2(V)};\ \inf_{V\in\mathbb{B}(1):\ \sigma^{-2}(V)>L} (L-L/2)\right\} \\ &\geqslant \min\left\{\inf_{V\in\mathbb{B}(1)} \frac{1}{2\sigma^2(V)};\ \frac{L}{2}\right\}, \end{aligned}$$

where the desired inverse inequality follows by letting $L\to+\infty$. $\square$

Proof of Theorem 2.4. (a) By Lemma 7.1, since $\mathscr{D}$ is dense in the unit ball of $L^\infty(\mathbb{R}^d)$ w.r.t. $\sigma(L^\infty, L^1)$,

$$\begin{aligned} &\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu(\|T_n - f\|_1 > r) \\ &\quad= \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu\left(\sup_{V\in\mathscr{D}} \langle T_n - f, V\rangle > r\right) \\ &\quad\geqslant \sup_{V\in\mathscr{D}} \liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}_\mu(\langle T_n - f, V\rangle > r) \\ &\quad\geqslant -\inf_{V\in\mathscr{D}} \inf\{J(g) | \langle g-f, V\rangle > r\} = -\inf\left\{J(g) |\ \sup_{V\in\mathscr{D}} \langle g-f, V\rangle > \right\} \\ &\quad= -\inf_{g:\|g-f\|_1>r} J(g) = -I(r). \end{aligned}$$

Thus (2.10) follows from Lemma 7.2. The second claim follows easily from (2.10) by means of the extra condition on T_n and H1 (as in Step 1 of the proof of Theorem 2.2).

(b) It follows from Theorem 2.3 and Lemma 7.2. $\square$

References

[1] D. Bosq, F. Merlevède, M. Peligrad, Asymptotic normality for density kernel estimators in discrete and continuous time, J. Multivariate Anal. 68 (1) (1999) 78–95.
[2] M. Carrasco, X. Chen, Mixing and moment properties of various GARCH and stochastic volatility models, Econometric Theory 18 (1) (2002) 17–39.
[3] M. Csörgö, L. Horváth, Central limit theorems for L_p-norms of density estimators, Z. Wahrsch. Verw. 80 (1988) 269–291.
[4] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, second ed., Springer, New York, 1998.
[5] K.D. Deuschel, D.W. Stroock, Large Deviations, Pure Applied Mathematics, vol. 137, Academic Press, San Diego, CA, 1989.
[6] L. Devroye, The equivalence of weak, strong and complete convergence in L^1 for kernel density estimates, Ann. Statist. 11 (3) (1983) 896–904.
[7] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, I–IV, Commun. Pure Appl. Math. 28 (1975) 1–47, 279–301; 29 (1976) 389–461; 36 (1983) 183–212.
[8] F.Q. Gao, Moderate deviations and large deviations for kernel density estimators, J. Theoret. Probab. 16 (2003) 401–418.
[9] F.Q. Gao, Moderate deviations and the law of iterated logarithm for kernel density estimators in L^1, preprint 2003.
[10] E. Giné, D.M. Mason, A.Y. Zaitsev, The L^1-norm density estimator process, Ann. Probab. 31 (2) (2003) 719–768.
[11] L. Horváth, On L_p-norms of multivariate density estimators, Ann. Statist. 19 (1991) 1933–1949.
[12] T. Kato, Perturbation Theory for Linear Operators, Springer, New York, 1992 (second corrected printing of the second edition).
[13] I. Kontoyiannis, S.P. Meyn, Spectral theory and limit theorems for geometrically ergodic Markov processes, Ann. Appl. Probab. 13 (1) (2003) 304–362.
[14] L. Lei, L. Wu, B. Xie, Large deviations and deviation inequality for kernel density estimator in $L_1(R^d)$-distance, in: Development of Modern Statistics and Related Topics, Series in Biostatistics vol. 1, World Scientific Press, Singapore, 2003, pp. 89–97.
[15] L. Lei, L. Wu, The exponential convergence of kernel density estimator in L^1 for ϕ-mixing processes, Ann. l'ISUP 48 (1–2) (2004) 59–68.
[16] D. Louani, Large deviations for the L_1-distance in kernel density estimation, J. Statist. Plann. Inference (2000) 177–182.
[17] S.P. Meyn, R.L. Tweedie, Markov Chains and Stochastic Stability, Springer, Berlin, 1993.
[18] M. Peligrad, Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption, Stochastics Stochastics Rep. 40 (1992) 147–168.
[19] E. Rio, Inégalités de Hoeffding pour les fonctions lipchitziennes de suites dépendantes, C.R. Acad. Sci. Paris Ser. I 330 (2000) 905–908.
[20] D. Ruelle, Thermodynamic Formalism, Addison-Wesley, Reading, MA, 1978.
[21] L.M. Wu, Moderate deviations of dependent random variables related to CLT, Ann. Probab. 23 (1995) 420–445.
[22] L.M. Wu, An introduction to large deviations, in: J.A. Yan, S. Peng, S. Fang, L. Wu (Eds.), Several Topics in Stochastic Analysis, Academic Press of China, Beijing, 1997, pp. 225–336 (in Chinese).
[23] L.M. Wu, Uniformly integrable operator and large deviations for Markov processes, J. Funct. Anal. 172 (2000) 301–376.