Bernoulli **12**(1), 2006, 65–83

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

LIANGZHEN LEI

[1]*Laboratoire de Mathématiques, CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France.*
E-mail: lei@math.univ-bpclermont.fr
[2]*Department of Mathematics, Wuhan University, 430072 Hubei, China*

We consider a reversible $\mathbb{R}^d$-valued Markov process $\{X_i;\ i \geqslant 0\}$ with the unique invariant measure $\mu(\mathrm{d}x) = f(x)\mathrm{d}x$, where the density f is unknown. The large-deviation principles for the nonparametric kernel density estimator f_n^* in $L^1(\mathbb{R}^d, \mathrm{d}x)$ and for $\|f_n^* - f\|_1$ are established. This generalizes the known results in the independent and identically distributed case. Furthermore, we show that f_n^* is asymptotically efficient in the Bahadur sense for estimating the unknown density f.

Keywords: Bahadur efficiency; kernel density estimator; large deviations; reversible Markov processes; uniformly integrable operators

1. Introduction

Let $\{X_n;\ n \geqslant 0\}$ be a reversible $\mathbb{R}^d$-valued Markov chain, defined on the probability space $(\Omega, (F_n^0)_{(n\in\mathbb{N})}, \mathcal{F}, (\mathbb{P}_x)_{x\in\mathbb{R}^d})$, with (unknown) Markov transition kernel $P(x, \mathrm{d}y)$. Assume that

(H1) P is irreducible (Meyn and Tweedie 1993) and symmetric with respect to the unique invariant probability measure μ, which is absolutely continuous, that is, $\mathrm{d}\mu(x) = f(x)\mathrm{d}x$, where the density f is unknown.

Given the observed sample $\{X_0, \ldots, X_n\}$, consider the empirical measure of the ladder type, that is,

$$L_n = \frac{1}{n}\left(\sum_{i=1}^{n-1} \delta_{X_i} + \frac{1}{2}(\delta_{X_0} + \delta_{X_n})\right) = \frac{1}{n}\sum_{i=0}^{n-1} \frac{1}{2}\left(\delta_{X_i} + \delta_{X_{i+1}}\right).$$

Let $K : \mathbb{R}^d \to R$ be a measurable function such that

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

and set

1350–7265 © 2006 ISI/BS

$$K_h(x) = \frac{1}{h^d} K\left(\frac{x}{h}\right)$$

for any $h > 0$. The kernel density estimator of the unknown function f is defined as follows: for all $x \in \mathbb{R}^d$,

$$f_n^*(x) := K_{h_n} * \mathrm{d}L_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} \frac{1}{2h_n^d}\left(K\left(\frac{x - X_i}{h_n}\right) + K\left(\frac{x - X_{i+1}}{h_n}\right)\right), \tag{1.2}$$

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

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

A natural distance of f_n^* from the unknown f is the L^1-distance,

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

The large-deviation behaviour of f_n^* in $(L^1(\mathbb{R}^d),\, \|\cdot\|_1 := \|\cdot\|_{L^1(\mathbb{R}^d)})$ is the subject of our study. In the independent and identically distributed (i.i.d.) case, due to Devroye (1983), all types of $L^1(\mathbb{R}^d)$-consistency of f_n^* are equivalent to condition (1.3) on the bandwidth. The asymptotic normality of D_n^* was investigated by Csörgő and Horváth (1988). Louani (2000) established the large-deviation principle (LDP) for D_n^*, and recently Lei *et al*. (2003) proved the weak LDP for f_n^* in $L^1(\mathbb{R}^d)$, and showed that the corresponding LDP is false. More recently Gao (2003) obtained the moderate deviation principle for f_n^* in $L^1(\mathbb{R}^d)$ and the law of the iterated logarithm for D_n^*. Giné *et al.* (2003) established a functional central limit theorem and a Glivenko–Cantelli theorem for the density estimator process in L^1-norm.

A natural question is how to extend those results from the i.i.d. case to the dependent case. Consistency of ar f_n^* has been studied by Peligrad (1992) and Bosq *et al*. (1999); see also the references therein. But little is known about large deviations. Large-deviation probabilities for f_n^* in L^1 and for D_n^* were obtained by Lei and Wu (2005) for uniformly ergodic Markov processes. Here uniform ergodicity means that there exist $1 \leqslant N \in \mathbb{N}^*$ and $C \geqslant 1$ such that

$$\frac{1}{C}\mu(\cdot) \leqslant \frac{1}{N}\sum_{k=1}^{N} P^k(x, \cdot) \leqslant C\mu(\cdot), \qquad \forall x \in E,$$

where E is a measurable subset of $\mathbb{R}^d$. The assumption is not satisfied by many discrete models with non-compact state space. For example, all real-valued stationary and ergodic Gaussian Markov processes are reversible but not uniformly ergodic. The purpose of this work is to establish the LDP for f_n^* in $L^1(\mathbb{R}^d)$ and for D_n^* in the framework of (H1) and (H2) below, instead of the strong 'uniform ergodicity' assumption.

(H2) For some $N \geqslant 1$, P^N is uniformly integrable in $L^2(\mu)$, that is, $\{(P^N f)^2;\ \|f\|_{L^2(\mu)} \leqslant 1\}$ is uniformly integrable.

Wu (2000a) proved that (H2) is a sufficient condition to obtain the LDP of L_n in the space $M_1(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with respect to the τ-topology (this condition is even necessary in the reversible case; see Wu 2002). The rate function is given by

$$J_\mu(\nu) := \begin{cases} \sup\{\int \log(u/Pu)\mathrm{d}\nu;\ 1 \leqslant u \in b\mathcal{B}\}, & \forall \nu \in M_1(\mathbb{R}^d),\ \nu \ll \mu, \\ +\infty, & \text{otherwise}, \end{cases} \tag{1.5}$$

where $b\mathcal{B}$ is the space of bounded and Borel-measurable functions on $\mathbb{R}^d$. We refer the reader to Wu (2000a) for related references on the subject.

This paper is organized as follows. The main results are stated in the next section. In Section 3 we present several crucial lemmas which may be of interest in their own right. We prove the main results in the rest of the paper.

2. Main results

In this paper, we use the following notation:

$$L^p := L^p(\mathbb{R}^d) := L^p(\mathbb{R}^d, \mathrm{d}x), \qquad \|f\|_p = \|f\|_{L^p(\mathbb{R}^d,\mathrm{d}x)}, \qquad L^p(\mu) := L^p(\mathbb{R}^d, \mu).$$

We denote, for any $L \geqslant 1$,

$$\mathcal{A}_{\mu,2}(L) := \left\{ \nu \in M_1(\mathbb{R}^d);\ \nu \ll \mu,\ \left\| \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \right\|_{L^2(\mu)} \leqslant L \right\}, \qquad \mathcal{A}_{\mu,2} := \bigcup_{L \geqslant 1} \mathcal{A}_{\mu,2}(L).$$

Throughout this paper we assume (H1) and (H2).

When the bandwidth $h_n \to 0$, $f_n^* \,\mathrm{d}x$ is 'close' to L_n in the τ-topology, and one may hope that $f_n^* \,\mathrm{d}x$ satisfies the same LDP as L_n. This intuition is in fact sound:

Theorem 2.1. *Assume $h_n \to 0$ (without* (1.3)*). Then $\mathbb{P}_\nu(f_n^* \in \cdot)$ satisfies, uniformly over initial measures $\nu \in \mathcal{A}_{\mu,2}(L)$ for each $L \geqslant 1$, the LDP in L^1 with respect to the weak topology $\sigma(L^1, L^\infty)$ with the rate function*

$$J(g) := \begin{cases} J(g\mathrm{d}x), & \textit{if } g\mathrm{d}x \in M_1(\mathbb{R}^d) \textit{ and } g\mathrm{d}x \ll f\mathrm{d}x; \\ +\infty, & \textit{otherwise}, \end{cases} \tag{2.1}$$

where $J(\cdot)$ is the Donsker–Varadhan entropy given in (1.5). *More precisely, J is inf-compact on $(L^1, \sigma(L^1, L^\infty))$, and for any measurable subset A of L^1, for every $L \leqslant 1$,*

$$-\inf_{g \in A^{\circ\sigma}} J(g) \leqslant \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(f_n^* \in A)$$

$$\leqslant \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu \in \mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(f_n^* \in A) \leqslant -\inf_{g \in \bar{A}^\sigma} J(g)$$

where $A^{\circ\sigma}$ and *$\bar{A}^\sigma$ denote respectively the interior and the closure of A with respect to the weak topology $\sigma(L^1, L^\infty)$.*

The LDP with respect to the weak topology on L^1 as above is too weak in the sense that it does not entail consistency, that is, $D_n^* \to 0$ in probability. As far as statistical issues are concerned, the main objects to be studied are:

(i) $\mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta)$, where $g\mathrm{d}x \in M_1(\mathbb{R}^d)$ is fixed, which is important in testing the hypothesis $H_0 : \mathrm{d}\mu(x) = f(x)\mathrm{d}x$ against $H_1 : \mathrm{d}\mu(x) = g(x)\mathrm{d}x$; or
(ii) $\mathbb{P}_\nu(D_n^* > \delta)$, whose statistical importance is obvious.

Unfortunately, Theorem 2.1 cannot be applied to these, since $\{\tilde{g} \in L^1; \|\tilde{g} - g\|_1 < \delta\}$ is not open in $\sigma(L^1, L^\infty)$ and $\{\tilde{g} \in L^1; \|\tilde{g} - f\|_1 \geqslant \delta\}$ is not closed in $\sigma(L^1, L^\infty)$. Therefore, in order to deal with objects (i) and (ii), we turn to Theorems 2.2 and 2.3 below.

Theorem 2.2. *Assume* (1.3). *Then for any* $L \geqslant 1$ *and for each* $\delta > 0$,

$$-I(\delta) \leqslant \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > \delta)$$

$$\leqslant \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 \geqslant \delta) \leqslant -I(\delta-), \tag{2.2}$$

where

$$I(\delta) = \inf\{J(g) | g \in L^1, \|g - f\|_1 > \delta\} > 0 \tag{2.3}$$

and $I(\delta-)$ *is the left limit of* I *at* δ.

Theorem 2.3. *Assume* (1.3). *Then* $\mathbb{P}_\nu(f_n^* \in \cdot)$ *satisfies the weak* LDP with rate function* J *on* $(L^1, \|\cdot\|_1)$ *uniformly over initial measures* $\nu \in \mathcal{A}_{\mu,2}(L)$ *for any* $L \geqslant 1$, *that is, for any* $L \geqslant 1$ *and* $g \in L^1$,

$$\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta)$$

$$= \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_1 < \delta) = -J(g). \tag{2.4}$$

With the above results, we have established the deviation estimates of the estimator f_n^*, which are useful in statistics. Now, we claim that f_n^* is asymptotically optimal in the Bahadur sense. Let Θ be the set of unknown data (P, μ) satisfying (H1) and (H2). Given a subset $\mathcal{D}$ of the unit ball in $b\mathcal{B}$, we say that an estimator $T_n(x) := T_n(x; X_0, \ldots, X_n) \in L^1(\mathbb{R}^d, \mathrm{d}x)$ is an *asymptotically* $\sigma(L^1, \mathcal{D})$*-consistent estimator* of the density f, if for all $V \in \mathcal{D}$, $\int_{\mathbb{R}^d} T_n(x)V(x)\mathrm{d}x \to \int_{\mathbb{R}^d} f(x)V(x)\mathrm{d}x$ in probability measure $\mathbb{P}_\mu$.

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

(a) *(Bahadur-type lower bound) Assume that* $\mathcal{D}$ *is dense in the unit ball of* L^∞ *with respect to the weak* topology* $\sigma(L^\infty, L^1)$. *Then for any* $\sigma(L^1, \mathcal{D})$*-asymptotically consistent estimator* T_n *of the unknown density* f,

$$\liminf_{r\to 0+}\frac{1}{r^2}\liminf_{n\to\infty}\frac{1}{n}\log \mathbb{P}_\mu(\|T_n - f\|_1 > r)$$

$$\geqslant -\frac{1}{2\sup_{\|V\|\leqslant 1}\sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}}\sigma^2(1_A)}, \tag{2.5}$$

where

$$\sigma^2(V) := 2\sum_{k=0}^{\infty}\langle V, P^k(V-\mu(V))\rangle_\mu - \mathrm{var}_\mu(V). \tag{2.6}$$

If, moreover, $\|T_n - T_n\circ\theta^N\|_1 \leqslant \delta_n \to 0$, *then* (2.5) *still holds with* $\mathbb{P}_\mu$ *substituted by* $\mathbb{P}_\nu$ *for any initial measure* $\nu \in M_1(E)$, *where* θ *is the shift on* Ω.

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

$$\liminf_{r\to 0+}\frac{1}{r^2}\liminf_{n\to\infty}\frac{1}{n}\log\inf_{\nu\in\mathcal{A}_{\mu,2}(L)}\mathbb{P}_\nu(\|f_n^* - f\|_1 > r)$$

$$= \limsup_{r\to 0+}\frac{1}{r^2}\limsup_{n\to\infty}\frac{1}{n}\log\sup_{\nu\in\mathcal{A}_{\mu,2}(L)}\mathbb{P}_\nu(\|f_n^* - f\|_1 > r)$$

$$= -\frac{1}{2\sup_{\|V\|\leqslant 1}\sigma^2(V)} = -\frac{1}{8\sup_{A\in\mathcal{B}}\sigma^2(1_A)}. \tag{2.7}$$

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 in the direction V of our statistical model Θ.

3. Preliminary lemmas

For every $V \in b\mathcal{B}$, put

$$P^V(x, \mathrm{d}y) := \exp\left(\frac{V(x)+V(y)}{2}\right)P(x, \mathrm{d}y).$$

We have the Feynman–Kac formula,

$$(P^V)^n f(x) = \mathbb{E}^{\mathbb{P}_x} f(X_n) exp\left(\sum_{k=0}^{n-1}\frac{V(X_k)+V(X_{k+1})}{2}\right),$$

where $\mathbb{E}^{\mathbb{P}_x}$ is the expectation with respect to $\mathbb{P}_x$. Introducing the Cramér functional

$$\Lambda^{(2)}(V) := \lim_{n\to\infty}\frac{1}{n}\log\|(P^V)^n\|_{L^2(\mu)\to L^2(\mu)}, \tag{3.1}$$

then $\mathrm{e}^{\Lambda^{(2)}(V)}$ is the spectral radius of P^V on $L^2(\mu)$. For the sake of convenience, we will write $\Lambda(V)$ for $\Lambda^{(2)}(V)$. It is well known (see Wu 2000a) that

$$J_\mu(\nu) = \sup\{\nu(V) - \Lambda(V) | V \in b\mathcal{B}\}, \qquad \forall \nu \in M_1(\mathbb{R}^d). \tag{3.2}$$

On the other hand, by the continuity of Λ on $b\mathcal{B}$ with respect to the Mackey topology proved in Wu (2000a, Theorem 5.1 and Theorem B.5) and by the Fenchel—Legendre theorem, we have, for all $t \in \mathbb{R}$,

$$\Lambda(t[V - \mu(V)]) = \sup\{\nu(tV) - t\mu(V) - J_\mu(\nu);\ \nu \in M_1(E)\} = \sup_{r\in\mathbb{R}}\{tr - J_V(r)\}, \tag{3.3}$$

where $J_V(r)$ is given by

$$J_V(r) := \inf\{J_\mu(\nu);\ \nu \in M_1(E),\ \ \nu(V) = \mu(V) + r\}. \tag{3.4}$$

J_V is convex. By the LDP of $\mathbb{P}_\nu(L_n \in \cdot)$ in Wu (2000a, Theorem 5.1) ($\nu \in \mathcal{A}_{\mu,2}$) and the contraction principle, $J_V : \mathbb{R} \to [0, +\infty]$ is inf-compact on $\mathbb{R}$ and $\mathbb{P}_\nu(L_n(V) - \mu(V) \in \cdot)$ satisfies the LDP with the rate function J_V. Furthermore, by the Fenchel–Legendre theorem and (3.3), we have

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

for $\Lambda(t[V - \mu(V)]) = \Lambda(tV) - t\mu(V)$. When $r \geqslant 0$, the supremum above can be taken only for $t \geqslant 0$. Then we obtain

$$J_V(r) = \begin{cases} \sup_{t\in\mathbb{R}}(t[r + \mu(V)] - \Lambda(tV)), & \forall r \in \mathbb{R}, \\ \sup_{t\geqslant 0}(t[r + \mu(V)] - \Lambda(tV)), & \forall r \geqslant 0. \end{cases} \tag{3.5}$$

Part (b) of the following lemma is crucial and gives us a robust estimate which extends the well-known inequality of Cramér in the i.i.d. case.

Lemma 3.1. *For the positive operator*

$$P^V(x, \mathrm{d}y) := \exp\left(\frac{V(x) + V(y)}{2}\right) P(x, \mathrm{d}y).$$

(a) *P^V is also symmetric in $L^2(\mu)$ and $\|P^V\|_{L^2(\mu)} = \mathrm{e}^{\Lambda(V)}$, and there exists $\phi \in L^2(\mu)$ μ-almost surely strictly positive such that $\int_E \phi^2 \mathrm{d}\mu = 1$ and*

$$P^V\phi = \mathrm{e}^{\Lambda(V)}\phi \qquad \text{over } \mathbb{R}^d,\ \mu\text{-a.s.}$$

Moreover, the eigenspace $Ker\left(\mathrm{e}^{\Lambda(V)} - P^V\right)$ of P^V associated with the eigenvalue $\mathrm{e}^{\Lambda(V)}$ in $L^2(\mu)$ is spanned by ϕ.

(b) *(A deviation inequality of Cramér type) For any initial measure $\nu \in \mathcal{A}_{\mu,2}$, $r > 0$,*

$$\mathbb{P}_\nu(L_n(V) > \mu(V) + r) \leqslant \mathrm{e}^{-nJ_V(r)} \cdot \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \tag{3.6}$$

where $J_V(r) = \inf\{J_\mu(\nu);\ \nu(V) = \mu(V) + r\}$.

(c) *Define a Markov kernel Q^V as*

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

Then $\nu_V := \phi^2\mu$ is the unique invariant probability measure for Q^V, and Q^V is symmetric on $L^2(\nu_V)$.

The Cramér type inequality (3.6) was established by Wu (2000b) in the continuous-time case.

Proof. (a) Under (H1) and (H2), P^V is again symmetric, uniformly integrable and irreducible on $L^2(\mu)$. Thus this part follows by Wu (2000a, Theorem 3.1 and Corollary 3.3).

(b) By the symmetry of P^V on $L^2(\mu)$, we have $\|(P^V)^n\|_{L^2(\mu)} := \|(P^V)^n\|_{L^2(\mu)\to L^2(\mu)} = \mathrm{e}^{n\Lambda(V)}$ for each $V \in b\mathcal{B}$. Thus for any initial measure $\nu \in \mathcal{A}_{\mu,2}$, $0 \leqslant f \in L^2(\mu)$ and any $t \in \mathbb{R}$,

$$E^\nu\left(f(X_n)\mathrm{e}^{ntL_n(V)}\right) \leqslant \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \|(P^{tV})^n\|_{L^2(\mu)}$$

$$= \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \mathrm{e}^{n\Lambda(tV)}. \tag{3.7}$$

By Chebychev's inequality,

$$\mathbb{E}^\nu\left(1_{[L_n(V)>\mu(V)+r]}f(X_n)\right) \leqslant \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \inf_{t\geqslant 0} \mathrm{e}^{-nt(\mu(V)+r)+n\Lambda(V)}$$

$$= \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \cdot \|f\|_{L^2(\mu)} \cdot \mathrm{e}^{-nJ_V(r)},$$

where the second equality follows from (3.5). So (3.6) holds.

(c) It is easy to verify that Q^V is a Markov kernel, that $\nu_V := \phi^2\mu$ is an invariant measure of Q^V, and that it is symmetric on $L^2(\nu_V)$. As Q^V is irreducible as well as P, $\phi^2\mu$ is the unique invariant measure of Q^V. □

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

Lemma 3.2. (a) $\Lambda(V)$ *is Gâteaux-differentiable on $b\mathcal{B}$.*

(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 (H2), $(P^V)^N$ is uniformly integrable on $L^2(\mu)$, and P^V is irreducible. Thus by Wu (2000a, Theorem 3.11), the largest eigenvalue $\mathrm{e}^{\Lambda(V)}$ of P^V is isolated in the spectrum $\sigma(P^V)$ of P^V on $L^2(\mu)$, with simple algebraic multiplicity. Consequently, by the theory of perturbation of linear operators (Kato 1984, Chapter VII, Theorem 1.8), $\mathrm{e}^{\Lambda(V)}$ is real-analytic on $b\mathcal{B}$, that is, $\Lambda(V + t\tilde{V})$ is analytic on $t \in \mathbb{R}$ for any V, $\tilde{V} \in b\mathcal{B}$ fixed.

(b) First of all, $\liminf_{n\to\infty} \Lambda(V_n) \geqslant \Lambda(V)$ by (3.3). The converse inequality which is equivalent to $\limsup_{n\to\infty} \mathrm{e}^{\Lambda(V_n)} \leqslant \mathrm{e}^{\Lambda(V)}$, follows by applying Wu (2000a, Proposition 3.8) to $\pi_n := (P^{V_n})^N$. □

Lemma 3.3 (Gibbs-type principle). *Given a function $V \in b\mathcal{B}$, a probability measure $\nu \ll \mu$ on $\mathbb{R}^d$ satisfies*

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

if and only if $\nu = \nu_V := \phi^2\mu$, where ϕ is the right eigenfunction of P^V associated with $\mathrm{e}^{\Lambda(V)}$ given in Lemma 3.1(a) satisfying $\mu(\phi^2) = 1$.

Proof. The proof is identical to that of Lei and Wu (2005, Lemma 3.4). □

Lemma 3.4. *Under* (H1) *and* (H2)*, for each $\nu = g\mathrm{d}x \in M_1(\mathbb{R}^d)$ satisfying $J_\mu(\nu) < +\infty$, there exists a sequence of $(\nu_{V_n} = \phi_n^2\mathrm{d}\mu)$ given in Lemma 3.1, such that*

$$\|\nu_{V_n} - \nu\|_{\mathrm{TV}} \to 0 \quad \textit{and} \quad \limsup_{n\to\infty} J(\nu_{V_n}) \leqslant J_\mu(\nu).$$

Here $\|\cdot\|_{\mathrm{TV}}$ means the total variation of a signed measure.

Proof. The proof is omitted; for details, we refer the reader to Lei and Wu (2005, Part 2, Proof of Theorem 2.2). □

Lemma 3.5. *Under* (H1) *and* (H2)*, we have the following:*

(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 any $V \in b\mathcal{B}$,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\Lambda(tV)|_{t=0} = \sigma^2(V),$$

which is given by (2.6).

(b) *Let J_V be defined as in* (3.4)*. Then J_V is strictly convex on $[J_V < +\infty]^0 = (a, b)$, where*

$$a = \lim_{t\to-\infty}\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(tV) - \mu(V), \qquad b = \lim_{t\to+\infty}\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(tV) - \mu(V)$$

(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) Under (H1) and (H2), by Wu (2000a, Theorem 3.11), 1 is an isolated point in the spectrum $\sigma(P)$ in $L^2(\mu)$ (i.e. there exists a spectral gap). We prove the lemma only in the case where $-1 \notin \sigma(P)$ (which corresponds to the aperiodicity of the irreducible chain). Otherwise, one may consider the periodic decomposition as in Lei and Wu (2005).

As in Wu (1995), we apply the analytical perturbation theory in Kato (1984). For each $z \in \mathbb{C}$, consider P^{zV} as an operator acting on the complexified space $L^2(E, \mu; \mathbb{C})$, which is analytical in z in the sense of Kato (1984). Then for any $\eta \in (0, \frac{1}{2})$ sufficiently small, there exist $\delta > 0$ and $C > 0$ such that, for all $V \in b\mathcal{B}$ with $\|V\| \leqslant 1$,

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

$$\|E(z, V)1 - 1\|_{L^2(\mu)} < \tfrac{1}{2}, \ \|(P^{zV})^n(I - E(z, V))\|_{L^2(\mu)} \leqslant C(1 - 2\eta)^n, \qquad \forall n;$$

(3) $z \to \lambda_{\max}(P^{zV})$ and $z \to E(z, V)f$ are analytic in z for $|z| \leqslant 2\delta$ (for each $f \in L^2(\mu)$);

where properties (1) and (2) follow from Kato (1984, Chapter IV, Theorem 3.16), and property (3) follows from Kato (1984, Chapter VII, Theorem 1.8).

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

Let

$$\Lambda_n(zV) := \frac{1}{n} \log\langle 1, (P^{zV})^n 1\rangle_\mu.$$

By properties (1) and (2) above, we have

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

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,

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

Applying the above estimates to $k = 2$, we obtain

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\Lambda_n(tV)|_{t=0} = \frac{1}{n}\mathbb{E}^\mu\left(\sum_{k=0}^{n-1} \frac{V(X_k) + V(X_{k+1})}{2} - n\mu(V)\right)^2$$

$$\to \mathrm{var}_{\mathbb{P}_\mu}(V(X_0)) + 2\sum_{n=1}^{\infty} \mathrm{cov}_{\mathbb{P}_\mu}(V(X_0), V(X_n)) = \sigma^2(V).$$

(b) All other properties of $J_V(r) = \sup_{t\in\mathbb{R}}(tr - \Lambda(t[V - \mu(V)]))$ are easy consequences of (3.5) and part (a) by elementary convex analysis. □

4. Proof of Theorem 2.1

The desired LDP of f_n^* in $(L^1, \sigma(L^1, L^\infty))$ is equivalent to the LDP of $f_n^*(x)\mathrm{d}x$ on $M_1(\mathbb{R}^d)$ with respect to the τ-topology $\sigma(M_1(\mathbb{R}^d), b\mathcal{B})$. We divide its proof into two parts.

4.1. Upper bound

By the abstract Gärtner–Ellis theorem in Wu (1997, p. 290, Theorem 2.7) and (3.2), it is enough to show that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in A_{\mu,2}(L)} \mathbb{E}^\nu \exp\left(n\int_{\mathbb{R}^d} f_n^*(y)V(y)\mathrm{d}y\right) \leqslant \Lambda(V), \tag{4.1}$$

and that $\Lambda(V)$ is monotonical continuous at 0, that is, if (V_n) is a sequence in $b\mathcal{B}$ decreasing pointwise to 0 over $\mathbb{R}^d$, then $\Lambda(V_n) \to 0$.

The second condition is satisfied by Lemma 3.2(b). It remains to verify (4.1). Put $V_n = (K_{h_n} * V)$; then $\|V_n\| \leqslant \|V\|$ and

$$n\int_{\mathbb{R}^d} f_n^*(y)V(y)\mathrm{d}y = \frac{1}{2}\sum_{k=0}^{n-1}(V_n(X_k) + V_n(X_{k+1})).$$

Consequently, we have for each $\nu \in \mathcal{A}_{\mu,2}(L)$,

$$\mathbb{E}^\nu \exp(n\int_{\mathbb{R}^d} f_n^*(y)V(y)\mathrm{d}y) \leqslant \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_{L^2(\mu)} \cdot \|(P^{V_n})^n\|_{L^2(\mu)} \leqslant L\cdot \mathrm{e}^{n\Lambda(V_n)}.$$

Thus

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in A_{\mu,2}(L)} \mathbb{E}^\nu \exp(n\int_{\mathbb{R}^d} f_n^*(y)V(y)\mathrm{d}y \leqslant \lim_{n\to\infty} \Lambda(V_n) = \Lambda(V)$$

where the last inequality follows from Lemma 3.2(b), for $V_n \to V$, $\mathrm{d}x$-almost everywhere. So (4.1) holds.

Remark 4.1. From the upper bound above, we can derive the following exponential convergence: for any $g_1, \ldots, g_m \in b\mathcal{B}$ and for any $\delta > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x\left(\max_{1\leqslant i\leqslant m}\left|\int_{\mathbb{R}^d}[f_n^*(x) - f(x)]g_i(x)\mathrm{d}x\right| \geqslant \delta\right) \\
&\leqslant -\inf\left\{J_\mu(g);\ \max_{1\leqslant i\leqslant m}\left|\int_{\mathbb{R}^d}[g(x) - f(x)]g_i(x)\mathrm{d}x\right| \geqslant \delta\right\} \\
&< 0, \qquad \mu - \text{a.s. } x \in E.
\end{aligned} \tag{4.2}$$

In fact, the last inequality follows from the inf-compactness of $J_\mu(\cdot)$ on $(M_1(E), \tau)$ and the fact that $J_\mu(\nu) = 0$ if and only if $\nu = \mu$ by (H1). For the first inequality, let $-c(\delta)$ be the non-positive constant on the right-hand side. For any $\varepsilon > 0$, using the proved upper bound, we have

$$\int_E \sum_{n=1}^{\infty} \mathbb{P}_x\left(\max_{1\leqslant i\leqslant m} \left| \int_{\mathbb{R}^d} [f_n^*(x) - f(x)] g_i(x) \mathrm{d}x \right| \geqslant \delta \right) \mathrm{e}^{-[c(\delta)+\varepsilon]n} \mu(\mathrm{d}x) < +\infty,$$

which yields

$$\sum_{n=1}^{\infty} \mathbb{P}_x\left(\max_{1\leqslant i\leqslant m} \left| \int_{\mathbb{R}^d} [f_n^*(x) - f(x)] g_i(x) \mathrm{d}x \right| \geqslant \delta \right) \mathrm{e}^{-(c(\delta)+\varepsilon))n} < +\infty, \ \mu\text{-a.s. } x.$$

Thus (4.2) holds μ-a.s. (for $\varepsilon > 0$ is arbitrary).

4.2. Lower bound

For the desired uniform lower bound, it is enough to prove that for any τ-neighbourhood $N(\nu, \delta) := \{\nu' \in M_1(\mathbb{R}^d);\ |(\nu' - \nu)(g_i)| < \delta,\ i = 1, \ldots, m\}$, $g_i \in b\mathcal{B}$ with $|g_i| \leqslant 1$, $\delta > 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(f_n^*(y)\mathrm{d}y \in N(\nu, \delta)) \geqslant -J_\mu(\nu), \ \mu\text{-a.s.} \tag{4.3}$$

(the arguments for this reduction, similar to those in the proof of Theorem 5.1 in Wu 2000a, are left to the reader). The proof of (4.3) is divided into two steps.

Step 1. The case $\nu = \nu_V$ for some $V \in b\mathcal{B}$. The idea of this step is borrowed from Donsker and Varadhan (1975a; 1975b; 1976; 1983). Given $V \in b\mathcal{B}$, let Q^V be the transition kernel defined in Lemma 3.1 and $\nu_V = \phi^2\mu$. By Lemma 3.1, Q^V is symmetric.

Let $\mathbb{Q}^V_{\omega(0)}$ be the law of the Markov process with transition kernel Q^V and starting point $\omega(0)$, which is ν_V-a.s. well defined on $\Omega = E^{\mathbb{N}}$, and $\mathbb{Q}^V := \int \mathbb{Q}^V_{\omega(0)} \mathrm{d}\nu_V(\omega(0))$. Denoting by $\xi(\omega)$ the density of $\mathbb{Q}^V_{\omega(0)}$ with respect to $\mathbb{P}_{\omega(0)}$ on $\sigma(X_1)$, we have for μ-a.s $\omega(0)$, on $\mathcal{F}_n := \sigma(X_k;\ 0 \leqslant k \leqslant n)$,

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

and $\mathbb{E}^{\mathbb{Q}^V} \log \xi = J^{(2)}(\mathbb{Q}^V|_{\mathcal{F}_1}) = J(\nu_V)$ by Lemma 3.3. For any $\varepsilon > 0$, setting

$$W_n := \left\{ \omega : \left| \int_{\mathbb{R}^d} g_i(x)[f_n^*(x, \omega) - \phi^2(x)]\mathrm{d}x \right| < \delta, \forall i = 1, \ldots, m \right\}$$

$$D_{n,\varepsilon} := \left\{ \omega : \frac{1}{n} \sum_{n=0}^{n-1} \log \xi(\theta^k \omega) \leqslant J(\nu_V) + \varepsilon \right\},$$

we have for μ-a.s. $\omega(0)$,

$$P_{\omega(0)}(W_n) \geqslant \int_{W_n} \exp\left(-\sum_{k=0}^{n-1} \log \xi(\theta^k \omega)\right) \mathrm{d}\mathbb{Q}^V_{\omega(0)}$$

$$\geqslant \exp[-n(J(\nu_V)+\varepsilon)] \cdot \mathbb{Q}^V_{\omega(0)}(W_n \cap D_{n,\varepsilon}). \tag{4.4}$$

So to obtain (4.3), it remains to show that $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$ and $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$, as n goes to infinity (for any $\varepsilon > 0$), for μ-a.s. $\omega(0)$.

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

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

which shows $\mathbb{Q}^V_{\omega(0)}(D_{n,\varepsilon}) \to 1$. To prove $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$, apply (4.2) in Remark 4.1 to Q^V (instead of P) which again satisfies (H1) and (H2); then

$$\mathbb{Q}^V_{\omega(0)}(W_n^c) \to 0, \ \nu_V\text{-a.s. } \omega(0).$$

The desired convergence holds.

Step 2. The general case. In order to prove (4.3) for general ν such that $J_\mu(\nu) < +\infty$, it is enough to approximate ν by ν_{V_n} as claimed in Lemma 3.4.

5. Proof of Theorem 2.2

The proof is divided into two parts.

Part 1. Lower bound in (2.2). The lower bound is an easy consequence of Theorem 2.1. Actually, as $\{g \in L^1; \|g - f\|_1 > \delta\}$ is open in the weak topology $\sigma(L^1, L^\infty)$, by Theorem 2.1, we have for any $L \geqslant 1$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{\nu \in A_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - f\|_1 > \delta) \geqslant - \inf_{g \in L^1; \|g-f\|_1 > \delta} J(g) = -I(\delta).$$

Part 2. Upper bound in (2.2). The proof of the upper bound is much more difficult, and it is divided into three steps, the first two similar to Devroye (1983) and the third inspired by Louani (2000).

Step 1. Approximation of K. As in Lei and Wu (2005), we may approximate K by

$$K^{(\varepsilon)} = \sum_{j=1}^{m} \lambda_j \frac{1_{A_j}}{|A_j|},$$

where $\int_{j=1}^m \lambda_\mathrm{j} = 1$, and A_j, $j \geqslant 1$, are disjoint finite rectangles in $\mathbb{R}^d$ of the form $\prod_{i=1}^d [x_i, x_i + a_i)$, so it is enough to establish (2.2) only for $K = 1_A/|A|$ where $A := \prod_{i=1}^d [x_i, x_i + a_i)$ (for details, see Lei and Wu 2005, step 1, part 2, proof of Theorem 2.3). Here $|A|$ denotes the Lebesgue measure of A.

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

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\big(\|f_n^* - K_{h_n} * f\|_1 > \delta\big) \leqslant -I(\delta-). \tag{5.1}$$

Note that

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

$$\leqslant \frac{1}{|A|h_n^d} \int |L_n(x+h_nA) - \mu(x+h_nA)|\mathrm{d}x.$$

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

Let $A^* = \Pi_{i=1}^d [x_i + 1/p, x_i + a_i - 1/p)$. 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,

$$\int |f_n^*(x) - K_{h_n} * f(x)|\mathrm{d}x$$

$$\leqslant \frac{1}{|A|h_n^d} \int \sum_{B\in\Psi, B\subseteq x+hA} |L_n(B) - \mu(B)|\mathrm{d}x + \frac{1}{|A|h_n^d} \int \{\mu(C_x) + L_n(C_x)\}\mathrm{d}x. \tag{5.2}$$

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

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

(using Fubini's theorem), the last term in (5.2) is bounded from above by

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

when p is large enough. 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 on the right-hand side of (5.2) from above by

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

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

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

By Lemma 3.1, we have for all $t > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\{L_n(S^c_{OR}) - \mu(S^c_{OR}) > \varepsilon\}$$

$$\leqslant -J_{1_{S^c_{OR}}}(\varepsilon) \leqslant -\Big(t[\varepsilon + \mu(S^c_{OR})] - \Lambda(t1_{S^c_{OR}})\Big).$$

Since $\lim_{R\to\infty} \Lambda(t1_{S^c_{OR}}) = 0$ by Lemma 3.2, for any $M > 0$, the left-hand side above is bounded from above by $-M$ for all R large enough, say $R \geqslant R_1$. Fix such $R \geqslant R_0 \vee R_1$ below. Summarizing these estimates, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\left(\int |f^*_n(x) - K_{h_n} * f(x)|\mathrm{d}x > \delta\right)$$

$$\leqslant (-M) \vee \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\left(\sum_{B\in\Psi, B\cap S_{OR}\neq\varnothing} |L_n(B) - \mu(B)| > \delta - 3\varepsilon\right). \tag{5.3}$$

Step 3. It remains to control the last term in (5.3). Set

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

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

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

where $\{-1, 1\}^{\tilde{\Psi}}$ denotes the set of all $\mathcal{B}(\tilde{\Psi})$-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 fixed $r > 0$,

$$\mathbb{P}_\nu\left(\sum_{B\in\Psi, B\cap S_{OR}\neq\varnothing} |L_n(B) - \mu(B)| > r\right) \leqslant \mathbb{P}_\nu\left(\max_{V\in\{-1,1\}^{\tilde{\Psi}}} L_n(V) - \mu(V) > r\right)$$

$$\leqslant \sum_{V\in\{-1,1\}^{\tilde{\Psi}}} \mathbb{P}_\nu(L_n(V) - \mu(V) > r).$$

By Lemma 3.1(b), for each $V \in \{-1, 1\}^{\tilde{\Psi}}$ and for all $r > 0$,

$$\sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(L_n(V) - \mu(V) > r) \leqslant \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \exp(-nJ_V(r)) \left\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right\|_2 \leqslant L\exp(-nJ_V(r)).$$

Secondly, the number of elements $\tilde{\Psi}$ is no greater than $(2Rp/h_n + 2)^d + 1 = o(n)$ by (1.3), and $\{-1, 1\}^{\tilde{\Psi}}$ has $2^{\#\tilde{\Psi}} = 2^{o(n)}$ elements for n large enough. Consequently, letting $\mathbb{B}(1)$ be the unit ball in $L^\infty(\mu)$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\left(\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset} |L_n(B) - \mu(B)| > r\right)$$

$$\leqslant \limsup_{n\to\infty} \frac{1}{n} \log 2^{o(n)} L \sup_{V\in\mathbb{B}(1)} \exp(-nJ_V(r)) = -\inf_{V\in\mathbb{B}(1)} J_V(r).$$

Combining with (5.3), we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu\left(\int |f_n^*(x) - K_{h_n} * f(x)|\mathrm{d}x > \delta\right)$$

$$\leqslant (-M) \vee \left(-\inf_{V\in\mathbb{B}(1)} J_V(\delta - 3\varepsilon)\right).$$

Since J_V is convex, non-decreasing and left-continuous on $[0, +\infty)$, using $\|\nu - \mu\|_{\mathrm{TV}} = \sup_{\|V\|\leqslant 1}[\nu(V) - \mu(V)] = 2\sup_{A\in\mathcal{B}}|\nu(A) - \mu(A)|$ and (3.4), we have

$$I(\delta) = \inf\left\{ J_\mu(\nu) \Big| \sup_{\|V\|\leqslant 1} [\nu(V) - \mu(V)] > \delta \right\}$$

$$= \inf_{\|V\|\leqslant 1} \inf_{r>\delta} J_V(r) = \inf_{V\leqslant 1} J_V(\delta+). \tag{5.4}$$

As $M > 0$ is arbitrary and $\lim_{\varepsilon\to 0+} \inf_{V\in\mathbb{B}(1)} J_V(\delta - 3\varepsilon) = I(\delta-)$ by (5.4), we obtain the desired (5.1) and then complete the proof of the upper bound in (2.2).

6. Proof of Theorem 2.3

The proof is divided into two parts, the first for the upper bound and the second for the lower bound.

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

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\nu\in\mathcal{A}_{\mu,2}(L)} \mathbb{P}_\nu(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) \leqslant -\inf_{\tilde{g};\|\tilde{g}-g\|_1\leqslant\delta} J(\tilde{g}).$$

Letting $\delta \to 0$, we obtain the desired result by the lower semi-continuity of J (which follows from (3.2)).

Part 2. Large-deviation lower bound. It is enough to prove that for all $g \in \mathcal{P}(E)$,

$$\lim_{\delta\to 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_x(\|f_n^* - g\|_{L^1(\mathbb{R}^d)} < \delta) = -J(g), \qquad \mu\text{-a.s.}$$

(This implies the desired uniform lower bound as in Wu 2000a.) The proof is divided into two steps.

Step 1. The $g\mathrm{d}x = \nu_V$ *case.* The proof of this case is parallel to that of Step 1 in the proof of (4.3): the only difference is that we now set

$$W_n := \{\omega : \|f_n^*(\omega) - g\|_1 < \delta\}$$

and the key point is to prove $\mathbb{Q}^V_{\omega(0)}(W_n) \to 1$, ν_V-a.s. Applying the upper bound in Theorem 2.2 to Q^V (instead of P), we have $\mathbb{Q}^V(W_n^c) \to 0$ at exponential rate. Using the Borel–Cantelli lemma, $\mathbb{Q}^V_{\omega(0)}(W_n^c) \to 0$ for ν_V-a.s. $\omega(0)$.

Step 2. The general case. To complete the proof, it remains to show the claim that for all $\nu = g\mathrm{d}x \in M_1(\mathbb{R}^d)$ satisfying $J(g) < +\infty$, there exists a sequence (ν_{V_n}) such that $\|\nu_{V_n} - \nu\|_{\mathrm{TV}} \to 0$, and $\limsup_{n\to\infty} J(\nu_{V_n}) \leqslant J_\mu(\nu)$. This was settled in Lemma 3.4.

7. Proof of Theorem 2.4

Lemma 7.1. *Let* $V \in b\mathcal{B}$. *If* T_n *is an asymptotically consistent estimator of* $\langle V, f\rangle := \int_E V(x)f(x)\mathrm{d}x$, *that is, for each* $(P, \mu) \in \Theta$ *(satisfying* (H1) *and* (H2)*),* $|\langle T_n, V\rangle - \langle f, V\rangle| \to 0$ *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 left-hand side of (7.1) is greater than $-J(g)$ for every $g \in \mathcal{P}$ which satisfies $\langle g - f\rangle, V > \delta$ and $J(g) < +\infty$. By step 2 in the proof of the lower bound of Theorem 2.1, it suffices to prove this in the case of $g = \nu_{\tilde V}$ where $\tilde V \in b\mathcal{B}$ is arbitrary. The proof, completely parallel to step 1 in the proof of the lower bound of Theorem 2.1, is based on the fact that $(Q^{\tilde V}, \nu_{\tilde V}) \in \Theta$ again, so it is omitted. □

Lemma 7.2. *Under* (H1) *and* (H2)*, let* $I(\cdot)$ *be defined as in* (2.3). *Then*

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

Proof. We only prove the first equality in (7.2) (the proof of the second is easy). By (5.4) and Lemma 3.5(b), for any $V \in b\mathcal{B}$ with $\|V\| \leqslant 1$,

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

so

$$\lim_{r\to 0+}\frac{I(r)}{r^2}\leqslant \frac{1}{2\sup_{\|V\|\leqslant 1}\sigma^2(V)}.$$

For the converse inequality, let $L>1$ be arbitrary but fixed. For any $\delta>0$ small enough, we have by Lemma 3.5,

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

Thus by Taylor's formula to order 3, we obtain for any $V\in\mathbb{B}(1)$ and $r\in(0,\delta]$,

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

$$\geqslant r^2\left(L\wedge\sigma^{-2}(V)-\frac{[L\wedge\sigma^{-2}(V)]^2\sigma^2(V)}{2}\right)-\frac{(Lr)^3}{6}\cdot C(L\delta),$$

where the last inequality is obtained by taking $t=r[L\wedge\sigma^{-2}(V)]$. Thus by (5.4),

$$\liminf_{r\to 0+}\frac{I(r)}{r^2}=\liminf_{r\to 0+}\inf_{V\in\mathbb{B}(1)}\frac{J_V(r)}{r^2}$$

$$\geqslant \min\left\{\inf_{V\in\mathbb{B}(1):\sigma^{-2}(V)\leqslant L}\frac{1}{2\sigma^2(V))};\inf_{V\in\mathbb{B}(1):\sigma^{-2}(V)>L}(L-L/2)\right\}$$

$$\geqslant \min\left\{\inf_{V\in\mathbb{B}(1)}\frac{1}{2\sigma^2(V)};\frac{L}{2}\right\}$$

where the desired converse inequality follows from letting $L\to+\infty$. □

With these two lemmas we are in a position to prove Theorem 2.4.

(a) By Lemma 7.1, since $\mathcal{D}$ is dense in the unit ball of L^∞ with respect to $\sigma(L^\infty, L^1)$,

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\mu(\|T_n-f\|_1>r)\geqslant \sup_{V\in\mathcal{D}}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}_\mu(\langle T_n-f, V\rangle>r)$$

$$\geqslant -\inf_{V\in\mathcal{D}}\inf\{J(g)|\langle g-f, V\rangle>r\}$$

$$=-\inf\left\{J(g)|\sup_{V\in\mathcal{D}}\langle g-f, V\rangle>r\right\}=-\inf_{g:\|g-f\|_1>r}J(g)=-I(r).$$

Thus (2.5) follows from Lemma 7.2. The second claim easily follows from (2.5) by means of the extra condition on T_n and (H1).

(b) This follows from Theorem 2.3 and Lemma 7.2.

Acknowledgements

I am very grateful to Professor L. Wu for his advice and guidance and to the referees for their careful comments and constructive suggestions.

References

Bosq, D., Merlevède, F. and Peligrad, M. (1999) Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.*, **68**, 78–95.

Csörgő, M. and Horváth, L. (1988) Central limit theorems for L_p-norms of density estimators. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **80**, 269–291.

Devroye, L. (1983) The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, **11**, 896–904.

Donsker, M.D. and Varadhan, S.R.S. (1975a) Asymptotic evaluation of certain Markov process expectations for large time, I. *Commun. Pure Appl. Math.*, **28**, 1–47.

Donsker, M.D. and Varadhan, S.R.S. (1975b) Asymptotic evaluation of certain Markov process expectations for large time, II. *Commun. Pure Appl. Math.*, **28**, 279-301.

Donsker, M.D. and Varadhan, S.R.S. (1976) Asymptotic evaluation of certain Markov process expectations for large time, III. *Commun. Pure Appl. Math.*, **29**, 389–461.

Donsker, M.D. and Varadhan, S.R.S. (1983) Asymptotic evaluation of certain Markov process expectations for large time, IV. *Commun. Pure Appl. Math.*, **36**, 183–212.

Giné, E., Mason, D.M. and Zaitsev, A.Yu. (2003) The L_1-norm density estimator process. *Ann. Probab.*, **31**, 719–768.

Gao, F.Q. (2003) Moderate deviations and the law of iterated logarithm for kernel density estimators in L_1. Preprint.

Kato, T. (1984) *Perturbation Theory for Linear Operators*, 2nd corr. printing of the 2nd edn. Berlin: Springer-Verlag.

Lei, L.Z. and Wu, L.M. (2005) Large deviations of kernel density estimator in $L_1(\mathbb{R}^d)$ for uniformly ergodic Markov processes. *Stochastic Process. Appl.*, **115**, 275–298.

Lei, L.Z., Wu, L.M. and Xie, B. (2003) Large deviations and deviation inequality for kernel density estimator in $L_1(\mathbb{R}^d)$-distance. In H. Zhang and J. Huang (eds), *Development of Modern Statistics and Related Topics*, pp. 89–97. River Edge, NJ: World Scientific.

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

Meyn, S.P. and Tweedie, R.L. (1993) *Markov Chains and Stochastic Stability.* London: Springer-Verlag.

Peligrad, M. (1992) Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption. *Stochastics Stochastics Rep.*, **40**, 147–168.

Wu, L.M. (1995) Moderate deviations of dependent random variables related to CLT. *Ann. Probab.*, **23**, 420–445.

Wu, L.M. (1997) An introduction to large deviations. In: J.A. Yan, S.G. Peng, S.Z. Fang and L.M. Wu (eds), *Several Topics in Stochastic Analysis*, pp. 225–336. Beijing: Academic Press of China (in Chinese).

Wu, L.M. (2000a) Uniformly integrable operator and large deviations for Markov processes. *J. Funct. Anal.*, **172**, 301–376.

Wu, L.M. (2000b) A deviation inequality for non-reversible Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **36**, 435–445.

Wu, L.M. (2002) Large deviations for reversible Markov chains. Preprint.

Received April 2004 and revised March 2005