Journal of Statistical Physics, Vol. 117, Nos. 5/6, December 2004 (© 2004)

Large Deviations for Probabilistic Cellular Automata II

A. Eizenberg[1] and Y. Kifer[2]

Received March 18 2004; accepted July 13 2004

We obtain an upper large deviations bound which shows that for some models of probabilistic cellular automata (which are far away from the product case) the lower large deviation bound derived in Eizenberg and Kifer *J. Stat. Phys.* **108**: 1255–1280 (2002) is sharp, and so the corresponding large deviations phenomena cannot be described via the traditional Donsker–Varadhan form of the action functional. For models which are close to the product case we derive approximate large deviations bounds using the Donsker–Varadhan functional for the product case.

KEY WORDS: Large deviations; cellular automata; Markov chains.

1. INTRODUCTION

Let X_t, $t \in \mathbb{Z}^+$, be a time homogeneous Markov chain with a compact metric phase space Γ. Denote by $M(\Gamma)$ the set of the probability Borel measures defined on Γ equipped with the weak topology. Consider the sequence of occupational measures

$$\zeta_T = \frac{1}{T}\sum_{t=0}^{T-1} \delta(X_t), \quad T \in \mathbb{Z}^+, \tag{1.1}$$

where $\delta(x)$ is the unit measure concentrated at a point $x \in \Gamma$. Due to the well-known results of Donsker and Varadhan (7–9), under certain conditions the asymptotic behavior of the occupational measures ζ_T obeys

[1]Jerusalem College of Engineering, Jerusalem, Israel; e-mail: alexe@math.huji.ac.il
[2]Institute of Mathematics, The Hebrew University, Jerusalem, Israel; e-mail: kifer@math.huji.ac.il

0022-4715/04/1200-0845/0

the large deviations principle with the action functional $I : M(\Gamma) \to [0, \infty]$ defined for any $\nu \in M(\Gamma)$ by the formula

$$I(\nu) = -\inf \left\{ \int_\Gamma \log \left(\frac{E_x f(X_1)}{f(x)} \right) \nu(dx) : f \in \mathcal{U}_1 \right\}, \tag{1.2}$$

where $\mathcal{U}_1$ is the set of positive continuous functions defined on Γ. More precisely, if the Markov chain X_t is a Feller processes, then for any closed with respect to the week topology subset K of $M(\Gamma)$,

$$\limsup_{T\to\infty} \frac{\ln P_x \{\zeta_T \in K\}}{T} \leqslant - \inf_{\nu \in K} I(\nu) \tag{1.3}$$

uniformly with respect to $x \in \Gamma$ (see, for instance, ref. 8). Moreover, it is known that under some additional conditions, such as, for instance, the existence of continuous densities for transition probabilities of corresponding Markov chains or, more generally, certain uniformity conditions formulated in ref. 10, or irreducibility conditions formulated in refs. 1 and 2, the following lower large deviations bound:

$$\liminf_{T\to\infty} \frac{\ln P_x \{\zeta_T \in U\}}{T} \geqslant - \inf_{\nu \in U} I(\nu) \tag{1.4}$$

holds true for any open with respect to the weak topology subset U of $M(\Gamma)$ and for each $x \in \Gamma$ (see also refs. 6, 10 and 16). Note that under the latter conditions, the lower and upper bounds have the same rate functionals so they are optimal for the corresponding class of processes, and, moreover, they are uniform with respect to $x \in \Gamma$, or at least independent of the initial distribution of the Markov chain.

However, as it was pointed out in ref. 14, such assumptions, frequently, are not satisfied for a large class of Markov chains, usually called probabilistic cellular automata (PCA), arising, for instance, as models describing large system of automata or some interacting particle systems (see, e.g., the earlier works on PCA such as refs. 19 and 20, as well, as later papers: refs. 4, 5, 17 and 18). Moreover, we have conjectured in ref. 14, that for some natural examples of PCA the uniform lower bounds (1.4) given by the Donsker–Varadhan action functional are not valid, while the upper bounds (1.3) are not optimal (although they are surely valid), and they can be improved by means of some alternative action functional depending on the initial distribution of the Markov chain. Our conjecture was supported only by some preliminary example (see Example 2 on page

1272 of ref. 14), where the Markov chain has an unique invariant measure and the Donsker–Varadhan lower estimates failed only due to certain degenerations.

Still, suggesting in ref. 14 the possibility of large deviations of a non Donsker–Varadhan type we kept in mind, mainly, situations where the Markov chain has more than one invariant measure. In such situations the following legitimate question arises. Let μ_0 and μ_1 be two different ergodic invariant measures of the corresponding PCA and $U(\mu_1)$ is some neighborhood of μ_1 such that $\mu_0 \notin U(\mu_1)$, then due to the ergodic theorem

$$\lim_{T\to\infty} P_{\mu_0}\{\zeta_T \in U(\mu_1)\} = 0, \tag{1.5}$$

where P_{μ_0} is the probability distribution in the path space of the Markov chain X_t with the initial distribution μ_0. We find it interesting to estimate the rate of convergence of $P_{\mu_0}\{\zeta_T \in U(\mu_1)\}$ to zero, in view of the fact that the Donsker–Varadhan type upper bound (1.3) becomes useless in this situation, since $I(\nu) = 0$ for any invariant measure ν of the Markov chain (see, for instance, ref. 8). This question is quite natural in the dynamical systems setup (which can be viewed as very degenerate Markov processes) where often there exist a lot of ergodic invariant measures (see ref. 16). In Section 2 we will formulate some general assumptions on Markov chains such that the rate of convergence of $P_{\mu_0}\{\zeta_T \in U(\mu_1)\}$ to zero is exponential and it depends on the initial distribution and formulate there our general results describing this phenomena. We will apply these results to PCA in Section 3 postponing their proof till Section 5. In Section 4 we provide a specific class of non-trivial examples of PCA (introduced by Wasserstein in ref. 20 for a different purpose), where lower bounds of the Donsker–Varadhan type are not valid. Moreover, we will show that for this model the lower bound of ref. 14 together with the upper bound derived here give the optimal large deviations estimates. On the other hand, in Section 6 we will describe a class of PCA, which includes, in particular, the direct product of finite Markov chains, where the large deviations can be fully described by means of the Donsker–Varadhan action functional defined in (1.2). Finally, in Section 7 we will obtain approximate large deviation estimates for Markov chains which are small perturbations of the product Markov chain considered in Section 6.

2. THE MAIN RESULTS: THE GENERAL CASE

All the results of the present paper are derived under the following basic assumption:

A0. The process X_t, $t \in \mathbb{Z}^+$, is a time homogeneous Markov chain on a phase space$(\Gamma, \mathfrak{B})$, where Γ is a compact metric space, and $\mathfrak{B}$ is the Borel σ-algebra of Γ.

However, to obtain our upper large deviations bounds we should confine ourselves to the Markov chains possessing the following special property:

A1. There exist a measure $\mu_0 \in \mathrm{M}(\Gamma)$ invariant with respect to the Markov chain X_t and a sequence of finite open partitions Λ_k of Γ such that $\mu_0(A) > 0$ for each $A \in \Lambda_k$, $k \geqslant 1$, and for each $\varepsilon > 0$ one can choose a sequence of positive integers $t(k, \varepsilon)$, $k \geqslant 1$ satisfying the following conditions:

$$\lim_{k \to \infty} \frac{t(k, \varepsilon)}{k} = 1$$

and for any integer n, $k \geqslant 1$, and for any sequence $f_i : \Gamma \to \mathbb{R}$, $0 \leqslant i \leqslant n$, of Λ_k-measurable functions,

$$E_{\mu_0}\left(\prod_{0 \leqslant i \leqslant n} f_i\left(X_{i\,t(k,\varepsilon)}\right)\right) \leqslant (1+\varepsilon)^n \prod_{0 \leqslant i \leqslant n} E_{\mu_0} f_i.$$

Remark 2.1. In Section 3 we will describe a class of PCA for which these assumptions hold true.

As it was pointed out in section, the purpose of this paper is to obtain certain large deviation estimates for the occupational measures ζ_T of the Markov chain X_t by means of the family of action functionals S_μ: $M(\Gamma) \to [0, \infty]$ introduced in ref. 14, where μ is the corresponding initial distribution of X_t . For readers' convenience we will provide the independent definition of S_μ here.

Recall (see ref. 11, Section 2.3), that for any two measures μ_1, $\mu_2 \in M(\Gamma)$ and each finite Borel partition $\Delta = \{Q_1, Q_2, \ldots, Q_n\}$ of Γ *the relative entropy of the partition* Δ *with measure* μ_1 *with respect to* μ_2 is defined by the formula

$$H_{\mu_1\|\mu_2}(\Delta) = \sum_{i=1}^{n} \mu_1(Q_i) \ln \frac{\mu_1(Q_i)}{\mu_2(Q_i)} \tag{2.1}$$

provided $\mu_1(Q_i) = 0$ whenever $\mu_2(Q_i) = 0$, and setting $H_{\mu_1\|\mu_2}(\Delta) = \infty$, otherwise . Now we can define

$$S_{\mu_2}(\mu_1) = \limsup_{k \to \infty} \frac{1}{k} H_{\mu_1\|\mu_2}(\Delta_k) \tag{2.2}$$

for any μ_1, $\mu_2 \in M(\Gamma)$.

Remark 2.2. Observe, that our definition, generally speaking, depends on the choice of the sequence of partitions Λ_k. We will return to the discussion of this important point later in this section.

The most general upper large deviations bound of the present paper is the following result whose proof will be given in Section 5.

Proposition 2.1. Suppose that a Markov chain X_t and a measure $\mu_0 \in M(\Gamma)$ satisfy the assumptions **A0** and **A1**. Then for each $\mu \in M(\Gamma)$ such that $S_{\mu_0}(\mu) < \infty$ and each $\varepsilon > 0$ there exists an open neighborhood $U(\mu, \varepsilon)$ of μ such that

$$\limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U(\mu, \varepsilon)\}}{T} \leqslant -(S_{\mu_0}(\mu) - \varepsilon).$$

Moreover, if $S_{\mu_0}(\mu) = \infty$, then for each $N > 0$ there exists an open neighborhood $U(\mu, N)$ of μ such that

$$\limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U(\mu, N)\}}{T} \leqslant -N.$$

Note. Let us emphasize the fact that μ is an arbitrary probability Borel measure, not necessarily invariant with respect to the Markov chain.

Corollary 2.2. Suppose that a Markov chain X_t and a measure $\mu_0 \in M(\Gamma)$ satisfy the assumptions **A0** and **A1**. Then for any closed with respect to the weak topology subset K of $M(\Gamma)$,

$$\limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in K\}}{T} \leqslant -\inf_{\nu\in K} S_{\mu_0}(\nu) \tag{2.3}$$

Proof. The statement follows from Proposition 2.1 in the standard way using the compactness argument. ∎

Remark 2.3. In order to show that the estimates provided by Proposition 2.1 and Corollary 2.2 have any significance, we must demonstrate some example of a Markov chain X_t and of measures μ_0, $\mu \in M(\Gamma)$ such that $S_{\mu_0}(\mu) > I(\mu)$, while this Markov chain and the initial distribution μ_0 satisfy the assumptions **A0** and **A1**. More specifically, as indicated in Section 1, our upper bounds could be especially useful when μ is an invariant measure with respect to the Markov chain X_t, since in this case $I(\mu)=0$, and, therefore, the Donsker–Varadhan type upper bounds are ineffective.

Therefore, we have to provide examples, where μ is an invariant measure with respect to X_t, while $S_{\mu_0}(\mu) > 0$. In Section 3 we exhibit a natural class of PCA, where this happens to be true. Yet, the most accomplished results are achieved when μ is an ergodic measure of the Markov chain X_t, since in this case we can apply our lower bounds provided in ref. 14.

Recall, that in ref. 14 we derived our lower bounds for a rather general class of Markov chains assuming the following conditions:

H1. There exists a sequence of finite open partitions Λ_k of Γ, $k \geqslant 1$, such that $\Lambda_k \prec \Lambda_{k+1}$ for each $k \geqslant 1$, and $\max_{A \in \Lambda_k} \operatorname{diam} A \to 0$ as $k \to \infty$ (in particular $\mathfrak{B}$ is the minimal σ -algebra generated by partitions Λ_k , $k \geqslant 1$);

H2. For any $k \geqslant 1$, $x \in \Gamma$, $B \in \Lambda_k$,

$$P\ (x\,,\ B) := P_x\ (X_1 \in B) > 0. \tag{2.4}$$

H3. For any $k \geqslant 1$, $A \in \Lambda_k$, $B \in \Lambda_{k+1}$, $x, y \in B$,

$$P\,(x, A) = P\,(y, A)\,. \tag{2.5}$$

Under this general framework we proved the following lower bound closely related to the subject of the present paper.

Proposition 2.3. Suppose that a Markov chain X_t satisfies the condition **A0** and the conditions **H1**–**H3**. Let μ be an ergodic invariant measure with respect to the chain X_t. Then for any initial distribution μ_0 and for any open with respect to the weak topology neighborhood U of μ we have

$$P_{\mu_0}\{\xi_T \in U\} \geqslant \exp(-(\,S_{\mu_0}(\mu) + \delta\,)\,T\,)$$

provided $T \geqslant T(\delta)$.

Proof. This is, actually, Corollary 3.2 of ref. 14. ∎

Remark 2.4. It was pointed out in ref. 14 that traditional PCA models considered, for instance, in refs. 4, 5, 15 and 18, satisfy our conditions **H1**–**H3** (and, clearly, the condition **A0**).

Remark 2.5. As we have observed in Remark 2.2, the value of the functional $S_{\mu_0}(\mu)$ depends, generally speaking, on the choice of the sequence Λ_k (while this sequence, clearly, could be chosen in a more than one way). Therefore, in order to provide the large deviations estimates of Propositions 2.1 and 2.2 in their strongest version, one should try to pick up the partitions Λ_k in such a way, that the both propositions hold simultaneously, bridging the gap between the upper and the lower bounds. Keeping in mind this observation, we will formulate our main result.

Introduce some metric on $M(\Gamma)$ generating the weak topology. Let $U_\delta(\mu)$ be the ball of radius $\delta > 0$ (with respect to this metric) centered at $\mu \in M(\Gamma)$.

Theorem 1. (the large deviations principle for the ergodic measures). Suppose that a Markov chain X_t, a measure $\mu_0 \in M(\Gamma)$, and a sequence of finite open partitions Λ_k of Γ, $k \geqslant 1$, satisfy the conditions **A0**, **A1** and **H1**–**H3**. Let μ be an ergodic invariant measure with respect to the chain X_t. Then

$$\lim_{\delta\to 0} \limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U_\delta(\mu)\}}{T} = \lim_{\delta\to 0} \liminf_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U_\delta(\mu)\}}{T} = -S_{\mu_0}(\mu).$$

Proof. The statement follows immediately from Propositions 2.1 and 2.3. ▮

3. APPLICATIONS TO PCA

In this section we will apply our general results to some traditional PCA models. In order to make the present paper more self-contained, recall the general model of PCA considered in Section 4 of ref. 14, where we followed the approach of refs. 4, 5, 15 and 18. Namely, let K be a finite set. Set $\Gamma = K^{\mathbb{Z}^d}$ for some $d \geqslant 1$. For any $\Phi \subset \Phi_1 \subset \mathbb{Z}^d$ let $\pi_\Phi : K^{\Phi_1} \to K^\Phi$ be the natural projection. For any $\varphi \in K^\Phi$ set

$$A_\Phi(\varphi) = \{\gamma \in \Gamma : \pi_\Phi(\gamma) = \varphi\}. \tag{3.1}$$

Clearly, for any finite $\Phi \subset \mathbb{Z}^d$ the subsets $\{A_\Phi(\varphi) : \varphi \in K^\Phi\}$ form a finite partition of Γ. Moreover, the family of sets $A_\Phi(\varphi)$ for all possible Φ and $\varphi \in K^\Phi$ serves as a sub-base for the standard product discrete topology on Γ metrizable in the usual way. Namely, for any $\underline{z} = (z_1, \ldots, z_d) \in \mathbb{Z}^d$ introduce the norm $\|\underline{z}\| = \max_{1\leqslant i\leqslant d} |z_i|$. Then the product topology on Γ is induced by the metric

$$\rho(\gamma, \gamma') = \sum_{\underline{z}\in\mathbb{Z}^d} 2^{-\|\underline{z}\|} \tilde{\rho}(\gamma(\underline{z}), \gamma'(\underline{z})), \quad \gamma, \gamma' \in \Gamma, \tag{3.2}$$

where $\widetilde{\rho}(s_1, s_2) = 0$ if $s_1 = s_2$, $\widetilde{\rho}(s_1, s_2) = 1$ otherwise, for any $s_1, s_2 \in K$ (considering a configuration $\gamma \in \Gamma$ as a function $\gamma : \mathbb{Z}^d \to K$). It is well known that Γ equipped with the metric ρ is a compact. We assume that a Γ-valued Markov chain X_t, $t \in \mathbb{Z}^+$, satisfies the following conditions:

C1. For any $\underline{z} \in \mathbb{Z}^d$ a finite neighborhood $N(\underline{z}) \subset \mathbb{Z}^d$ of $\underline{z}$ is defined together with a local transition kernel $P^{\underline{z}} : K^{N(\underline{z})} \times K \to (0, 1)$. More precisely, for any $\eta \in K^{N(\underline{z})}$ a probability distribution $P^{\underline{z}}(\eta, \cdot)$ is defined on K. Recall that the elements of $N(\underline{z})$ are called *the neighbors* of $\underline{z} \in \mathbb{Z}^d$;

C2. The transition probability kernel $P(\cdot, \cdot)$ of X_t has the following property: for each $x \in \Gamma$, each finite $\Phi \subset \mathbb{Z}^d$, and each $\varphi \in K^{\Phi}$,

$$P(x,\ A_\Phi(\varphi)) = \prod_{\underline{z} \in \Phi} P^{\underline{z}}(\pi_{N(\underline{z})}(x), \varphi(\underline{z})). \tag{3.3}$$

C3. There exists an integer $n_0 \geqslant 1$ such that for each $\underline{z} \in \mathbb{Z}^d$,

$$N(\underline{z}) \subseteq \left\{ \underline{z}' \in \mathbb{Z}^d : \|\underline{z} - \underline{z}'\| \leqslant n_0 \right\}. \tag{3.4}$$

Remark 3.1. Notice, that the condition **C1** includes, in particular, the fact that $P^{\underline{z}}(\eta, \underline{z}) > 0$ for any $\eta \in K^{N(\underline{z})}$, $s \in K$. Clearly, this fact, together with **C2** yields $P(x, A_\Phi(\varphi)) > 0$ for any $x \in \Gamma$, each finite $\Phi \subset \mathbb{Z}^d$, and every $\varphi \in K^{\Phi}$.

Remark 3.2. It has been shown in ref. 14 that PCA model satisfying the conditions **C1–C3** automatically satisfies the conditions **A0** and **H1–H3** formulated in Section 2 of the present paper provided the sequence of partitions Λ_k is properly chosen, which can be done in a natural way. Now we are going to show that one can choose the sequence Λ_k, $k \geqslant 1$, in such a way that the assumptions **A0** and **A1** are satisfied, provided the condition **C3** is replaced by some more restrictive condition **C4** (see below in this section). However, as we have pointed out in Remark 2.5, in order to obtain the optimal value of the functional $S_{\mu_1}(\mu_2)$, that is, to meet the conditions of our Theorem 1, one should choose the sequence Λ_k, $k \geqslant 1$, in such a way that the assumptions **A0** and **A1** are satisfied simultaneously with **H1–H3**. We are going to show that such a choice is possible for some classes of PCA. Still, we will have to choose our sequence of partitions in a slightly more complicated way, than in ref. 14.

Since our main purpose is to illustrate some concrete application of our general results, we will restrict our consideration to the case when $d = 1$, that is, $\Gamma = K^{\mathbb{Z}}$. Consequently, we should adapt our notations to this case.

Notations. For $\gamma \in \Gamma$, $z \in \mathbb{Z}$ denote by γ_z the z coordinate of γ. For any given $a \leqslant b \in \mathbb{Z}$ denote $[a, b] = \{z \in \mathbb{Z}: a \leqslant z \leqslant b\}$ and $\Gamma_{[a,b]} = K^{[a,b]}$, then $\pi_{[a,b]}$ is the natural projection from Γ to $\Gamma_{[a,b]}$. Similarly, if $[a, b] \subset [a_1, b_1]$ where $a_1 \leqslant b_1 \in \mathbb{Z}$, we will denote by $\pi_{[a,b]}$ the natural projection from $\Gamma_{[a_1,b_1]}$ to $\Gamma_{[a,b]}$. More precisely, for each $\gamma = (\dots, \gamma_{-1}, \gamma_0, \gamma_1, \dots) \in \Gamma$ or $\gamma = (\gamma_{a_1}, \gamma_{a_1+1}, \dots, \gamma_{b_1}) \in \Gamma_{[a_1,b_1]}$ one has

$$\pi_{[a,b]}(\gamma) = (\gamma_a, \gamma_{a+1}, \dots, \gamma_b).$$

For the sake of simplicity, we will also use the notation $\pi_a = \pi_{[a,a]}$. Finally, similarly to (3.1), we will use the notations

$$A_{[a,b]}(\varphi) = \left\{\gamma \in \Gamma : \pi_{[a,b]}(\gamma) = \varphi\right\} \tag{3.5}$$

for any given $\varphi \in \Gamma_{[a,b]}$, $a \leqslant b \in \mathbb{Z}$ and

$$A_a(s) = \{\gamma \in \Gamma : \pi_a(\gamma) = s\}$$

for any given $s \in K$, $a \in \mathbb{Z}$.

Remark 3.3. Let $\mathfrak{B}$ be the Borel σ-algebra of Γ. Clearly, the family of sets $A_{[a,b]}(\varphi)$ for all possible $\varphi \in \Gamma_{[a,b]}$, $a \leqslant b \in \mathbb{Z}$ generates $\mathfrak{B}$.

The main condition of this section is

C4. The phase space $\Gamma = K^{\mathbb{Z}}$, that is $d = 1$. Moreover, there exists an integer $n_0 \geqslant 1$ such that for each $z \in \mathbb{Z}$,

$$N(z) = [z+1, z+n_0]. \tag{3.6}$$

Remark 3.4. Observe, that the condition **C4** yields **C3.** A simple class of PCA satisfying the conditions **C1, C2** and **C4,** with $K = \{0, 1\}$, $n_0 = 1$, was introduced by Wasserstein in ref. 20 for a different purpose. Later, it was pointed out by Föllmer in ref. 15, that, in some cases, Wasserstein's example permits infinitely many invariant measures. This example plays an important role in this paper and it will be described in details in the next section 4.

Now we reformulate the conditions **C1, C2** and **C4** in the form more suitable for our purpose. Let $d = 1$, $K = \{s_1, s_2, \dots, s_{m_0}\}$, then **C1**, actually, means that for any $z\in\mathbb{Z}$, $\eta\in K^{N(z)}$ we are given a finite sequence of numbers $P^z(\eta,\, s_m)>0$, $1\leqslant m\leqslant m_0$, such that $\sum_{1\leqslant m\leqslant m_0} P^z(\eta,\, s_m)=1$. Moreover, by **C2** and **C4,** one has

$$\begin{aligned} P(\gamma,\, A_{[a,\,b]}(\varphi)) &= P_\gamma\left\{\pi_{[a,\,b]}(X_1)=\varphi\right\} \\ &= \prod_{z\in[a,\,b]} P^z\left(\pi_{[z+1,\; z+n_0]}(\gamma),\, \varphi_z\right) \end{aligned} \tag{3.7}$$

for any $\gamma\in\Gamma$, $\varphi=(\varphi_a,\varphi_{a+1},\ldots,\varphi_b)\in\Gamma_{[a,\,b]}$, $a,\, b\in\mathbb{Z}$. In particular, when $a=b$, the formula (3.7) takes the form

$$P_\gamma\left\{\pi_a(X_1)=s\right\}=P^a\left(\pi_{[a+1,\,a+n_0]}(\gamma),\, s\right) \tag{3.8}$$

for any $\gamma\in\Gamma$, $s\in K$.

Remark 3.5. Due to Remark 3.3, the Markov chain X_t is, actually, completely defined by the formula (3.7), and, moreover, this formula is equivalent to the conditions **C1, C2** and **C4.**

Notice, that by (3.7), for all $a,\, b\in\mathbb{Z}$, for each $\varphi\in\Gamma_{[a,\,b]}$, $\eta\in\Gamma_{[a+1,\,b+n_0]}$, and for any $\gamma,\tilde{\gamma}\in A_{[a+1,\,b+n_0]}(\eta)$,

$$P(\gamma,\, A_{[a,\,b]}(\varphi))=P(\tilde{\gamma},\, A_{[a,\,b]}(\varphi)), \tag{3.9}$$

which enables us to introduce the notation

$$P(\eta,\, A_{[a,\,b]}(\varphi))=P(\gamma,\, A_{[a,\,b]}(\varphi)) \tag{3.10}$$

for each $\eta\in\Gamma_{[a+1,\,b+n_0]}$, provided $\gamma\in A_{[a+1,\,b+n_0]}(\eta)$.

The main purpose of this section is to show that the Markov chain X_t satisfying the conditions **C1, C2** and **C4** automatically satisfies the conditions of Section 2 (provided the corresponding sequence of partitions is properly chosen), which will allow us to reformulate our general results for the specific situation described in the present section. However, first we should prove some auxiliary result concerning our chain X_t, which is, actually, a generalization of the formula (3.9).

Proposition 3.1. Let X_t be a Markov chain satisfying the conditions **C1, C2** and **C4.** Let $a,\, b\in\mathbb{Z}$, $t\in\mathbb{Z}^+$, and $\varphi\in\Gamma_{[a,\,b]}$, $\eta\in\Gamma_{[a+t,\,b+t\,n_0]}$,. Then for any $\gamma,\tilde{\gamma}\in A_{[a+t,\,b+t\,n_0]}(\eta)$,

$$P_\gamma\left\{\pi_{[a,\,b]}(X_t)=\varphi\right\}=P_{\tilde{\gamma}}\left\{\pi_{[a,\,b]}(X_t)=\varphi\right\}. \tag{3.11}$$

Proof. For $t=1$ the formula (3.11) coincides with (3.9). Now we can complete the proof by induction in $t \in \mathbb{Z}^+$. Indeed, suppose that the statement is true for some $t \in \mathbb{Z}^+$and for any $a', b' \in \mathbb{Z}$, each $\varphi \in \Gamma_{[a', b']}$, $\eta \in \Gamma_{[a'+t, b'+t n_0]}$ and every $\gamma, \tilde{\gamma} \in A_{[a'+t, b'+t n_0]}(\eta)$. Then for any given $a, b \in \mathbb{Z}$, $\varphi \in \Gamma_{[a, b]}$, $\eta \in \Gamma_{[a+1+t, b+n_0+t n_0]}$ and all $\gamma, \tilde{\gamma} \in A_{[a+1+t, b+(1+t) n_0]}(\eta)$ by the Markov property and the notation (3.10) we obtain, setting $a' = a+1$, $b' = b+n_0$, that

$$\begin{aligned}
&P_\gamma \left\{\pi_{[a, b]} \left(X_{t+1}\right) = \varphi\right\} \\
&\quad = \sum_{\varphi' \in \Gamma_{[a+1, b+n_0]}} P_\gamma \left\{\pi_{[a+1, b+n_0]}(X_t) = \varphi'\right\} P(\varphi', A_{[a, b]}(\varphi)) \\
&\quad = \sum_{\varphi' \in \Gamma_{[a+1, b+n_0]}} P_{\tilde{\gamma}} \left\{\pi_{[a+1, b+n_0]}(X_t) = \varphi'\right\} P(\varphi', A_{[a, b]}(\varphi)) \\
&\quad = P_{\tilde{\gamma}} \left\{\pi_{[a, b]} \left(X_{t+1}\right) = \varphi\right\}. \quad \blacksquare
\end{aligned}$$

Remark 3.6. Due to the last assertion, we can introduce the following convenient notation :

$$P^{[a, b]}(\eta, \varphi, t) = P_\gamma \left\{\pi_{[a, b]}(X_t) = \varphi\right\} \tag{3.12}$$

provided $\gamma \in A_{[a+t, b+t n_0]}(\eta)$, where $\varphi \in \Gamma_{[a, b]}$, $\eta \in \Gamma_{[a+t, b+t n_0]}$, $a, b \in \mathbb{Z}$, and $t \in \mathbb{Z}^+$.

Proposition 3.1 yields that in order to meet the condition **A1**, it is enough to pick up an invariant measure $\mu_0 \in M(\Gamma)$ such that it is *ψ-mixing* (see ref. 3), or, more precisely, it satisfies the following conditions: for any $a \leqslant b \in \mathbb{Z}$, $\varphi \in \Gamma_{[a, b]}$,

$$\mu_0(A_{[a, b]}(\varphi)) > 0 \tag{3.13}$$

and for each $\varepsilon > 0$ there exists an integer $m(\varepsilon) > 0$ large enough such that if Φ_1, Φ_2 are finite subsets of $\mathbb{Z}$, $\operatorname{dist}(\Phi_1, \Phi_2) \geqslant m(\varepsilon)$, and $\varphi_1 \in K^{\Phi_1}$, $\varphi_2 \in K^{\Phi_2}$, then

$$\left| \frac{\mu_0(A_{\Phi_1}(\varphi_1) \cap A_{\Phi_2}(\varphi_2))}{\mu_0(A_{\Phi_1}(\varphi_1))\, \mu_0(A_{\Phi_2}(\varphi_2))} - 1 \right| \leqslant \varepsilon. \tag{3.14}$$

Notation. Denote by $M_{\psi i}(\Gamma)$ the set of all the measures $\mu \in M(\Gamma)$ invariant with respect to the Markov chain X_t and satisfying the conditions (3.13) and (3.14).

Remark 3.7. Note, that we are talking above about space ψ*-mixing* though we are interested about large deviations of the Markov chain X_t in time. Observe, also that if the Markov chain X_t satisfies the conditions **C1, C2** and **C4** and $\mu \in M(\Gamma)$ is invariant with respect to X_t, then by (3.7) for any $a \leqslant b \in \mathbb{Z}$, $\varphi \in \Gamma_{[a,b]}$,

$$\begin{aligned}
&\mu_0\left(A_{[a,b]}(\varphi)\right) \\
&\quad = \sum_{\eta \in \Gamma_{[a+1,\, b+n_0]}} \mu_0\left(A_{[a+1,\, b+n_0]}(\eta)\right) P\left(\eta,\ A_{[a,b]}(\varphi)\right) \\
&\quad \geqslant \min_{\eta \in \Gamma_{[a+1,\, b+n_0]}} P\left(\eta,\ A_{[a,b]}(\varphi)\right) \geqslant \alpha_0^{b-a+1} > 0,
\end{aligned}$$

where $\alpha_0 = \min_{\eta \in \Gamma_{[z+1,\, z+n_0]},\, a \leqslant z \leqslant b,\, s \in K} P^z(\eta, s)$. Therefore, the condition (3.13) is automatically satisfied. Moreover, one can easily verify that the similar fact is true when **C4** is replaced by **C3,** even when $d > 0$.

Now we are going to establish the following important property of measures belonging to the class $M_{\psi i}(\Gamma)$.

Proposition 3.2. Let X_t be a Markov chain satisfying the conditions **C1–C3** (with $d = 1$). Then each measure $\mu \in M_{\psi i}(\Gamma)$ is ergodic with respect to X_t.

Proof. Let $\Gamma = \Gamma_1 \cup \Gamma_2$, where Γ_1, Γ_2 are two nonempty disjoint Borel measurable subsets of Γ invariant with respect to the kernel $P(\cdot, \cdot)$, i.e., for each $\gamma \in \Gamma_i$, $i = 1, 2$,

$$P(\gamma,\ \Gamma_i) = 1. \tag{3.15}$$

Our goal is to prove that for each $\mu \in M_{\psi i}(\Gamma)$ either $\mu(\Gamma_1) = 0$ or $\mu(\Gamma_1) = 1$. We will divide the proof into three steps.

Step 1. Choose some $\gamma = (\ldots, \gamma_{-1}, \gamma_0, \gamma_1, \ldots) \in \Gamma_1$ and let $\eta \in \Gamma$ be such that $\eta_z = \gamma_z$ for each integer index z provided $|z| \geqslant N$ for some $N \geqslant 0$. We claim that $\eta \in \Gamma_1$. Indeed, for each cylindrical subset of the form $A_{[-m,m]}(\varphi)$, and by **C2** and **C3** for any given $\varphi \in \Gamma_{[-m,m]}$, $m > N + n_0$,

$$\begin{aligned}
\frac{P\left(\gamma,\ A_{[-m,m]}(\varphi)\right)}{P\left(\eta\, A_{[-m,m]}(\varphi)\right)} &= \prod_{z \in [-m,m]} \frac{P^z\left(\pi_{N(z)}(\gamma),\ \varphi_z\right)}{P^z\left(\pi_{N(z)}(\eta),\ \varphi_z\right)} \\
&= \prod_{z \in [-N-n_0,\, N+n_0]} \frac{P^z\left(\pi_{N(z)}(\gamma),\ \varphi_z\right)}{P^z\left(\pi_{N(z)}(\eta),\ \varphi_z\right)}
\end{aligned}$$

since $\pi_{N(z)}(\gamma) = \pi_{N(z)}(\eta)$, provided $|z| > N + n_0$.

Therefore,

$$C_N^{-1} \leqslant \frac{P(\gamma,\ A_{[-m,m]}(\varphi))}{P(\eta\, A_{[-m,m]}(\varphi))} \leqslant C_N\,, \tag{3.16}$$

where

$$C_N = \left(\max_{z\in[-N-n_0,\, N+n_0]}\ \max_{\zeta,\zeta'\in N(z),\, s\in K} \frac{P^z(\zeta,\ s)}{P^z(\zeta',\ s)}\right)^{2N+2n_0+1} > 0.$$

Since the estimate (3.16) holds for each cylindrical subset $A_{[-m,m]}(\varphi)$ such that $\varphi \in \Gamma_{[-m,m]}$ and $m > N + n_0$, and since these subsets generate the Borel σ-algebra $\mathfrak{B}$ of Γ, it follows that:

$$C_N^{-1} \leqslant \frac{P(\gamma,\ A)}{P(\eta\, A)} \leqslant C_N$$

for each $A \in \mathfrak{B}$. In particular,

$$C_N^{-1} \leqslant \frac{P(\gamma,\ \Gamma_2)}{P(\eta, \Gamma_2)} \leqslant C_N\,,$$

which together with (3.15) yields $P(\eta,\ \Gamma_2) = 0$, and, therefore, using (3.15) again, $\eta \in \Gamma_1$.

Step 2. For a given $n > 0$ denote by $\widetilde{\mathfrak{B}}_n$ the sub-σ-algebra of $\mathfrak{B}$ generated by the family of functions $\pi_z(\cdot)$, $|z| \geqslant n+1$. Let $m > 0$, and let $\varphi \in \Gamma_{[-m,m]}$. We claim that for each $\varepsilon > 0$ and for each $K \in \widetilde{\mathfrak{B}}_{m+m(\varepsilon)}$,

$$\begin{aligned}
&\mu(A_{[-m,m]}(\varphi))\mu(K)(1-\varepsilon)\\
&\leqslant \mu(A_{[-m,m]}(\varphi) \cap K)\\
&\leqslant \mu(A_{[-m,m]}(\varphi))\mu(K)(1+\varepsilon),
\end{aligned} \tag{3.17}$$

where $m(\varepsilon)$ has been defined in (3.14). Clearly, it is enough to prove (3.17) for each set $K \in \widetilde{\mathfrak{B}}_{m+m(\varepsilon)}$ of the form $K = A_{[m+m(\varepsilon),b]}(\xi_1) \cap A_{[-b,-m-m(\varepsilon)]}(\xi_2)$, where $b > m + m(\varepsilon)$, and $\xi_1 \in \Gamma_{[m+m(\varepsilon),b]}$, $\xi_2 \in \Gamma_{[-b,-m-m(\varepsilon)]}$. However, for such sets the claim follows immediately from (3.14).

Step 3. Now we will show that $\Gamma_1 \in \widetilde{\mathfrak{B}}_n$ for each $n > 0$. Indeed, for a given $n > 0$ fix some $\varphi_0 \in \Gamma_{[-n\,n]}$ and define a function $F : \Gamma \to A_{[-n,n]}(\varphi_0)$ in the following way: for each $\gamma \in \Gamma$ let $\eta = F(\gamma)$ be the unique $\eta \in A_{[-n,n]}(\varphi_0)$ such that $\eta_z = \pi_z(\gamma)$ for each integer index z provided

$|z|\geqslant n+1$. Clearly, F is a $\widetilde{\mathfrak{B}}_n$-measurable map, since its value is completely defined by the values of the functions $\pi_z(\cdot)$, $|z|\geqslant n+1$. Therefore, for each $A\in\mathfrak{B}$ we have $F^{-1}(A)\in\widetilde{\mathfrak{B}}_n$. However, by the definition of F, and according to the conclusion of *Step 1*, $\gamma\in\Gamma_1$ if and only if $F(\gamma)\in\Gamma_1\cap A_{[-n,n]}$. Thus, $F^{-1}(\Gamma_1)=\Gamma_1$, and, therefore, $\Gamma_1\in\widetilde{\mathfrak{B}}_n$. Furthermore, by (3.17), for each $m>0$, any $\varphi\in\Gamma_{[-m,m]}$, and every $\varepsilon>0$ we have

$$\begin{aligned}&\mu(A_{[-m,m]}(\varphi))\mu(\Gamma_1)(1-\varepsilon)\\ &\quad\leqslant\mu(A_{[-m,m]}(\varphi)\cap\Gamma_1)\\ &\quad\leqslant\mu(A_{[-m,m]}(\varphi))\mu(\Gamma_1)(1+\varepsilon).\end{aligned}$$

However, since we can chose $\varepsilon>0$ arbitrary small, this inequality yields, for each $m>0$ and each $\varphi\in\Gamma_{[-m,m]}$,

$$\mu(A_{[-m,m]}(\varphi))\mu(\Gamma_1)=\mu(A_{[-m,m]}(\varphi)\cap\Gamma_1),$$

which, actually, means that for each m the event Γ_1 is independent of the algebra generated by the partition $A_{[-m,m]}(\varphi)$, $\varphi\in\Gamma_{[-m,m]}$. Hence, Γ_1 is independent of the σ-algebra generated by all cylinder sets to which it itself belongs, and so $\mu(\Gamma_1)=(\mu(\Gamma_1))^2$ which sais that $\mu(\Gamma_1)=1$ or $=0$. ▮

Remark 3.8. In Section 4 we will demonstrate an important class of examples where the set $M_{\psi i}(\Gamma)$ is not empty. It is plausible that $M_{\psi i}(\Gamma)\neq\emptyset$ also in a more general situation.

Conjecture. Let X_t be a Markov chain satisfying the conditions **C1, C2** and **C4.** Then the set $M_{\psi i}(\Gamma)$is not empty.

Now we are going to derive some properties of measures from $M_{\psi i}(\Gamma)$ assuming that **C1, C2** and **C4** are satisfied, which will later enable us to use the results of Section 2.

Proposition 3.3. Let X_t be a Markov chain satisfying the conditions **C1, C2** and **C4**, and let $\mu_0\in M_{\psi i}$. Let $a\leqslant b\in\mathbb{Z}$, $\varepsilon>0$, and denote $t_\varepsilon=b-a+m(\varepsilon)$, where $m(\varepsilon)$ has been introduced in (3.14). Then for each integer $c\geqslant a$, each integer $n\geqslant 1$, and any sequence $\varphi_0,\varphi_1,\dots,\varphi_{n-1}\in\Gamma_{[a,b]}$, $\varphi_n\in\Gamma_{[a,c]}$,

$$P_{\mu_0}\left(\bigcap_{0\leqslant i\leqslant n-1}\left\{\pi_{[a,b]}\left(X_{i\,t_\varepsilon}\right)=\varphi_i\right\}\bigcap\left\{\pi_{[a,c]}\left(X_{nt_\varepsilon}\right)=\varphi_n\right\}\right)$$
$$\leqslant(1+\varepsilon)^n\,\mu_0\left(A_{[a,c]}(\varphi_n)\right)\prod_{0\leqslant i\leqslant n-1}\mu_0\left(A_{[a,b]}(\varphi_i)\right). \tag{3.18}$$

Proof. Since μ_0 is an invariant measure of the Markov chain X_t, we have, using the notation (3.12),

$$\begin{aligned}
&\mu_0\left(A_{[a,b]}(\varphi)\right)\\
&\quad=P_{\mu_0}\left\{\pi_{[a,b]}(X_t)=\varphi\right\}\\
&\quad=\int_\Gamma P_\gamma\left\{\pi_{[a,b]}(X_t)=\varphi\right\}\mu_0(d\gamma)\\
&\quad=\sum_{\eta\in\Gamma_{[a+t,\,b+t\,n_0]}}\mu_0\left(A_{[a+t,\,b+t\,n_0]}(\eta)\right)P^{[a,b]}(\eta,\varphi,t)
\end{aligned} \tag{3.19}$$

for any $a, b\in\mathbb{Z}$, for each $t\in\mathbb{Z}^+$ and for every $\varphi\in\Gamma_{[a,b]}$.

Next, observe, that by the condition (3.14), for each $\varepsilon>0$ and for any given $a_1\leqslant b_1<a_2\leqslant b_2\in\mathbb{Z}$, satisfying $a_2-b_1\geqslant m(\varepsilon)$ and any $\varphi_1,\in\Gamma_{[a_1,b_1]}, \varphi_2\in\Gamma_{[a_2,b_2]}$,

$$\mu_0\left(A_{[a_1,b_1]}(\varphi_1)\cap A_{[a_2,b_2]}(\varphi_2)\right)\leqslant(1+\varepsilon)\mu_0\left(A_{[a_1,b_1]}(\varphi_1)\right)\mu_0\left(A_{[a_2,b_2]}(\varphi_2)\right). \tag{3.20}$$

Now we will prove the proposition by induction in $n\geqslant 1$. First, we demonstrate the proof for $n=1$. In this case by (3.11) and the Markov property the left part of (3.18) takes the form

$$\begin{aligned}
&P_{\mu_0}\Big(\left\{\pi_{[a,b]}(X_0)\right.\\
&\quad\left.=\varphi_0\right\}\bigcap\left\{\pi_{[a,c]}\left(X_{t_\varepsilon}\right)=\varphi_1\right\}\Big)\\
&\quad=\int_{A_{[a,b]}(\varphi_0)}P_\gamma\left\{\pi_{[a,c]}\left(X_{t_\varepsilon}\right)=\varphi_1\right\}\mu_0(d\gamma)\\
&\quad=\sum_{\eta\in\Gamma_{[a+t_\varepsilon,\,c+t_\varepsilon n_0]}}\mu_0\left(A_{[a\,b]}(\varphi_0)\cap A_{[a+t_\varepsilon,\,c+t_\varepsilon n_0]}(\eta)\right)\\
&\qquad\quad\times P^{[a,c]}(\eta,\varphi,t_\varepsilon)
\end{aligned} \tag{3.21}$$

for any given $\varphi_0\in\Gamma_{[a,b]}$, $\varphi_1\in\Gamma_{[a,c]}$ (where we use the notation (3.12)). However, since $a+t_\varepsilon-b=m(\varepsilon)$, we have, by (3.20), that for any $\eta\in$

$\Gamma_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}$,

$$\mu_0\left(A_{[a\, b]}(\varphi_0)\cap A_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}(\eta)\right) \leqslant (1+\varepsilon)\mu_0\left(A_{[a\, b]}(\varphi_0)\right)\mu_0\left(A_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}(\eta)\right). \tag{3.22}$$

Therefore, by (3.21), (3.22) and (3.19),

$$\begin{aligned} &P_{\mu_0}\left(\left\{\pi_{[a,b]}(X_0)=\varphi_0\right\}\bigcap\left\{\pi_{[a,c]}\left(X_{t_\varepsilon}\right)=\varphi_1\right\}\right)\\ &\quad\leqslant(1+\varepsilon)\mu_0\left(A_{[a\, b]}(\varphi_0)\right)\\ &\qquad\times\sum_{\eta\in\Gamma_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}}\mu_0\left(A_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}(\eta)\right)P^{[a,c]}(\eta,\varphi_1,t_\varepsilon)\\ &\quad=(1+\varepsilon)\mu_0\left(A_{[a\, b]}(\varphi_0)\right)\mu_0\left(A_{[a,c]}(\varphi_1)\right), \end{aligned}$$

proving the proposition for the case $n=1$.
Now, suppose that the proposition is true for some $n\geqslant 1$, for each integer $c'\geqslant a$ and for any sequence $\varphi'_0, \varphi'_1, \dots, \varphi'_{n-1}\in\Gamma_{[a,b]}$, $\varphi'_n\in\Gamma_{[a,c']}$. Let $c\geqslant a$ $(c\in\mathbb{Z})$, and let $\varphi_0,\varphi_1,\dots,\varphi_n\in\Gamma_{[a,b]}$, $\varphi_{n+1}\in\Gamma_{[a,c]}$, then by (3.11) and the Markov property

$$\begin{aligned} &P_{\mu_0}\left(\bigcap_{0\leqslant i\leqslant n}\left\{\pi_{[a,b]}\left(X_{it_\varepsilon}\right)=\varphi_i\right\}\bigcap\left\{\pi_{[a,c]}\left(X_{(n+1)t_\varepsilon}\right)=\varphi_{n+1}\right\}\right)\\ &\quad=\sum_{\eta\in\Gamma_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}}P_{\mu_0}\left(A^n_{[a,b]}(\varphi_0,\dots,\varphi_n)\cap\left\{\pi_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}\left(X_{nt_\varepsilon}\right)=\eta\right\}\right)\\ &\qquad\times P^{[a,c]}(\eta,\varphi_{n+1},t_\varepsilon), \end{aligned} \tag{3.23}$$

where we set, for a given sequence $\varphi_0,\varphi_1,\dots,\varphi_n\in\Gamma_{[a,b]}$,

$$A^n_{[a,b]}(\varphi_0,\dots,\varphi_n)=\bigcap_{0\leqslant i\leqslant n}\left\{\pi_{[a,b]}\left(X_{it_\varepsilon}\right)=\varphi_i\right\}.$$

Next, for each given $\eta\in\Gamma_{[a+t_\varepsilon,\, c+t_\varepsilon n_0]}$ define the set

$$\begin{aligned} \Gamma(\eta,\varphi_n)=\{\tilde\eta=&\left(\tilde\eta_a,\tilde\eta_{a+1},\dots,\tilde\eta_{c+t_\varepsilon n_0}\right)\in\Gamma_{[a,\, c+t_\varepsilon n_0]}:(\tilde\eta_a,\dots,\tilde\eta_b)\\ &=\varphi_n,\left(\tilde\eta_{a+t_\varepsilon},\dots,\tilde\eta_{c+t_\varepsilon n_0}\right)=\eta\}, \end{aligned}$$

which is not empty since $a+t_\varepsilon\geqslant b+1$.

Now, assuming that the proposition is true for a given $n \geqslant 1$ and for each integer $c' \geqslant a$, we can write setting $c' = c + t_\varepsilon n_0$,

$$
\begin{aligned}
&P_{\mu_0}\left(A^n_{[a,b]}(\varphi_0, \dots, \varphi_n) \cap \left\{\pi_{[a,c]}\left(X_{n t_\varepsilon}\right) = \eta\right\}\right) \\
&\quad = \sum_{\tilde{\eta} \in \Gamma(\eta, \varphi_n)} P_{\mu_0}\left(\bigcap_{0 \leqslant i \leqslant n-1} \left\{\pi_{[a,b]}\left(X_{i t_\varepsilon}\right) = \varphi_i\right\} \cap \left\{\pi_{[a, c+t_\varepsilon n_0]}\left(X_{n t_0}\right) = \tilde{\eta}\right\}\right) \\
&\quad \leqslant (1+\varepsilon)^n \sum_{\tilde{\eta} \in \Gamma(\eta, \varphi_n)} \mu_0\left(A_{[a, c+t_\varepsilon n_0]}(\tilde{\eta})\right) \prod_{0 \leqslant i \leqslant n-1} \mu_0\left(A_{[a,b]}(\varphi_i)\right) \\
&\quad = (1+\varepsilon)^n \prod_{0 \leqslant i \leqslant n-1} \mu_0\left(A_{[a,b]}(\varphi_i)\right) \mu_0\left(A_{[a,b]}(\varphi_n) \cap A_{[a+t_\varepsilon, c+t_\varepsilon n_0]}(\eta)\right).
\end{aligned}
\tag{3.24}
$$

However, using again the definition of t_ε together with the formula (3.20), it follows:

$$
\begin{aligned}
&\mu_0\left(A_{[a,b]}(\varphi_n) \cap A_{[a+t_\varepsilon, c+t_\varepsilon n_0]}(\eta)\right) \\
&\quad \leqslant (1+\varepsilon)\, \mu_0\left(A_{[a,b]}(\varphi_n)\right) \mu_0\left(A_{[a+t_\varepsilon, c+t_\varepsilon n_0]}(\eta)\right).
\end{aligned}
\tag{3.25}
$$

Bringing together the formulas (3.23)–(3.25), and, finally, using (3.19), we obtain

$$
\begin{aligned}
&P_{\mu_0}\left(\bigcap_{0 \leqslant i \leqslant n} \left\{\pi_{[a,b]}\left(X_{i t_\varepsilon}\right) = \varphi_i\right\} \bigcap \left\{\pi_{[a,c]}\left(X_{(n+1) t_\varepsilon}\right) = \varphi_{n+1}\right\}\right) \\
&\quad \leqslant (1+\varepsilon)^{n+1} \mu_0\left(A_{[a,b]}(\varphi_n)\right) \prod_{0 \leqslant i \leqslant n-1} \mu_0\left(A_{[a,b]}(\varphi_i)\right) \\
&\qquad \times \sum_{\eta \in \Gamma_{[a+t_\varepsilon, c+t_\varepsilon n_0]}} \mu_0\left(A_{[a+t_\varepsilon, c+t_\varepsilon n_0]}(\eta)\right) P^{[a,c]}(\eta, \varphi_{n+1}, t_\varepsilon) \\
&\quad = (1+\varepsilon)^{n+1} \mu_0\left(A_{[a,c]}\left(\varphi_{n+1}\right)\right) \prod_{0 \leqslant i \leqslant n} \mu_0\left(A_{[a,b]}(\varphi_i)\right),
\end{aligned}
$$

completing the proof. ∎

Observe that, for technical reasons, the last proposition is formulated in a slightly more general way, than it is really needed. It is more convenient for our purpose to reformulate this proposition in the following final form.

Corollary 3.4. Let $a \leqslant b \in \mathbb{Z}$, $\varepsilon > 0$, and denote $t_\varepsilon = b - a + m(\varepsilon)$, where $m(\varepsilon)$ has been introduced in (3.14). Denote by $\mathfrak{B}_{[a,b]}$ the σ-algebra generated by the partition $\{A_{[a,b]}(\varphi) : \varphi \in \Gamma_{[a,b]}\}$. Then for each integer $n \geqslant 1$, and for any sequence $f_i : \Gamma \to \mathbb{R}$, $0 \leqslant i \leqslant n$, of $\mathfrak{B}_{[a,b]}$-measurable random variables,

$$E_{\mu_0}\left(\prod_{0\leqslant i\leqslant n} f_i\left(X_{i\,t_\varepsilon}\right)\right) \leqslant (1+\varepsilon)^n \prod_{0\leqslant i\leqslant n} E_{\mu_0} f_i.$$

Proof. Setting $c = b$ in the last proposition, we obtain for each integer $n \geqslant 1$, and any sequence $\varphi_0, \varphi_1, \ldots, \varphi_n \in \Gamma_{[a,b]}$,

$$P_{\mu_0}\left(\bigcap_{0\leqslant i\leqslant n} \left\{\pi_{[a,b]}\left(X_{i\,t_\varepsilon}\right) = \varphi_i\right\}\right) \leqslant (1+\varepsilon)^n \prod_{0\leqslant i\leqslant n} \mu_0\left(A_{[a,b]}(\varphi_i)\right), \quad (3.26)$$

which yields the assertion of the corollary. ∎

Next, in order to meet the conditions of Section 2, one should introduce a proper sequence of partitions. Namely, choose Λ_k, $k \geqslant 1$, to be the partition of Γ generated by the sets $A_{[-Q(k),k]}(\varphi)$ for all possible $\varphi \in \Gamma_{[-Q(k),k]}$ (with $A_{[a,b]}(\varphi)$ and $\Gamma_{[a,b]}$ introduced at the beginning of this section) and let

$$\lim_{k\to\infty} \frac{Q(k)}{k} = 0. \quad (3.27)$$

Now we can formulate the main result of this section.

Theorem 2. Let X_t be a Markov chain satisfying the conditions **C1, C2** and **C4**, and let $\mu_0 \in M_{\psi i}$. Let Λ_k, $k \geqslant 1$, be the sequence of partitions of Γ introduced above. Then

(a) The conditions **A0** and **A1** are satisfied. Moreover, for any closed with respect to the weak topology subset W of $M(\Gamma)$,

$$\limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in W\}}{T} \leqslant -\inf_{\nu\in W} S_{\mu_0}(\nu). \quad (3.28)$$

(b) Suppose, in addition, that $n_0 = 1$, and $\lim_{k\to\infty} Q(k) = \infty$, then the conditions **H1–H3** are also satisfied. Moreover, let μ be an ergodic measure with respect to the Markov chain X_t, and let $U_\delta(\mu)$ be the ball of radius $\delta > 0$ (with respect to some metric on $M(\Gamma)$ corresponding to the weak topology) centered at $\mu \in M(\Gamma)$. Then

$$\lim_{\delta\to 0} \limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U_\delta(\mu)\}}{T} = \lim_{\delta\to 0} \liminf_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U_\delta(\mu)\}}{T} = -S_{\mu_0}(\mu). \tag{3.29}$$

Proof. (a) Clearly, $\Gamma = K^{\mathbb{Z}}$ is a compact space with respect to the standard product topology which can be introduced by the metric (3.2) which for $d = 1$ has the form

$$\rho(\gamma, \gamma') = \sum_{z\in\mathbb{Z}} 2^{-|z|} \widetilde{\rho}(\gamma_z, \gamma'_z) \tag{3.30}$$

for any $\gamma, \gamma' \in \Gamma$, where $\widetilde{\rho}(s_1, s_2) = 0$ if $s_1 = s_2$, and $\widetilde{\rho}(s_1, s_2) = 1$ otherwise. Thus, **A0** is trivially satisfied. Next, substituting $a = -Q(k)$, $b = k$ in Corollary 3.4, and setting $t(k, \varepsilon) = b - a + m(\varepsilon) = k + Q(k) + m(\varepsilon)$ for each given $\varepsilon > 0$ (where $Q(k)$satisfies (3.27)), we obtain for all integers $n, k \geqslant 1$, and for any sequence $f_i : \Gamma \to \mathbb{R}$, $0 \leqslant i \leqslant n$ of Λ_k-measurable functions the following formula:

$$E_{\mu_0}\left(\prod_{0\leqslant i\leqslant n} f_i\left(X_{i\,t(k,\varepsilon)}\right)\right) \leqslant (1+\varepsilon)^n \prod_{0\leqslant i\leqslant n} E_{\mu_0} f_i,$$

which together with the fact that $\lim_{k\to\infty} t(k, \varepsilon)/k = 1$ yields **A1**. Now (3.28) follows by Corollary 2.2.

(b) If $\lim_{k\to\infty} Q(k) = \infty$, then by (3.30), the sequence of partitions Λ_k, $k \geqslant 1$, satisfies the condition **H1**. On the other hand, if $n_0 = 1$, the formula (3.9) yields **H3**. By **Remark 3.1**, the condition **H2** is also satisfied. We can apply, therefore, Theorem 1, obtaining (3.29). ▮

4. THE MAIN EXAMPLE

Next, we are going to describe a specific example of the situation when the lower bounds of the Donsker–Varadhan type cannot hold true.

We adapt for this purpose an example introduced by Wasserstein and, latter, considered by Föllmer for different purposes (see also Remark 3.5). Suppose that we are given a sequence of real numbers $\frac{1}{2} < p_z < 1$, where $z \in \mathbb{Z}$, then we can define *the local transition kernels* P^z for any $z \in \mathbb{Z}$, $s, s' \in \{0, 1\}$ by

$$P^z(s, s') = \begin{cases} p_z & \text{if } s = s', \\ 1 - p_z & \text{otherwise.} \end{cases} \tag{4.1}$$

Define a Markov chain X_t evolving on a phase space $\Gamma = \{0, 1\}^{\mathbb{Z}}$ with the transition probability kernel $P(\cdot, \cdot)$ given by the formula

$$P(\gamma, A_{[a,b]}(\varphi)) = P_\gamma \left\{ \pi_{[a,b]}(X_1) = \varphi \right\} = \prod_{z \in [a,b]} P^z \left(\gamma_{z+1}, \varphi_z \right) \tag{4.2}$$

for any $\gamma \in \Gamma$, $\varphi = (\varphi_a, \varphi_{a+1}, \ldots, \varphi_b) \in \Gamma_{[a,b]}$, $a, b \in \mathbb{Z}$. Clearly, the conditions **C1**, **C2** and **C4** are satisfied with $n_0 = 1$, and, moreover, the formula (4.2) is a particular case of (3.7).

The simplest particular case of measures from $M_{\psi i}(\Gamma)$ are the product invariant measures of X_t. More precisely, recall, that $\mu \in M(\Gamma)$ is called *a product measure* if for any $\varphi = (\varphi_a, \varphi_{a+1}, \ldots, \varphi_b) \in \Gamma_{[a,b]}$, $a \leqslant b \in \mathbb{Z}$,

$$\mu \{ A_{[a,b]}(\varphi) \} = \prod_{z \in [a,b]} \mu \{ A_z(\varphi_z) \}. \tag{4.3}$$

Denote by $M_{pi}(\Gamma)$ the set of all the product measures $\mu \in M(\Gamma)$ invariant with respect to the Markov chain X_t. The family $M_{pi}(\Gamma)$ was described in ref. 15 in the following way. Let $0 \leqslant r \leqslant 1$. Denote for any $z \in \mathbb{Z}$, $s \in \{0, 1\}$,

$$D_r^z(s) = \begin{cases} \frac{1}{2} + \left(r - \frac{1}{2} \right) d_z & \text{if } s = 1, \\ \frac{1}{2} - \left(r - \frac{1}{2} \right) d_z & \text{if } s = 0, \end{cases}$$

where $d_z = \prod_{i \geqslant z} (2p_i - 1)$. Obviously, $D_r^z(s) \neq 0$. Introduce the product measure $\nu_r \in M(\Gamma)$ by the formula

$$\nu_r \{ A_{[a,b]}(\varphi) \} = \prod_{z \in [a,b]} D_r^z(\varphi_z) \tag{4.4}$$

for any $\varphi = (\varphi_a, \varphi_{a+1}, \ldots, \varphi_b) \in \Gamma_{[a,b]}$, $a, b \in \mathbb{Z}$.

It is easy to verify by a direct computation that each measure defined by (4.4) belongs to the set $M_{pi}(\Gamma)$, and, moreover, $M_{pi}(\Gamma)$ consists of all the measures ν_r , $0 \leqslant r \leqslant 1$.

Assume that

$$d_0 = \prod_{i \geqslant 0} (2\, p_i - 1) > 0 \tag{4.5}$$

and let $0 \leqslant r_0 \neq r_1 \leqslant 1$, then our condition (4.5) yields $\nu_{r_0} \neq \nu_{r_1}$. In particular, it means that the transitional kernel (4.2) has infinitely many invariant measures. Now, as in Section 3, choose Λ_k, $k \geqslant 1$, to be the partition of Γ generated by the sets $A_{[-Q(k),k]}(\varphi)$ for all possible $\varphi \in \Gamma_{[-Q(k),k]}$, where

$$\lim_{k \to \infty} Q(k) = \infty, \qquad \lim_{k \to \infty} \frac{Q(k)}{k} = 0 \tag{4.6}$$

(*say,* $Q(k) = [\sqrt{k}]$ *for each integer* $k \geqslant 1$). The main result of this section is the following.

Proposition 4.1. Let X_t be the Markov chain defined by (4.2) such that the condition (4.5) is satisfied, Λ_k, $k \geqslant 1$, be the sequence of partitions of Γ introduced in the previous paragraph, and $\mu_0 \in M_{pi}(\Gamma)$. Then

(a) the conditions **A0**, **A1**, and **H1–H3** are satisfied,

(b) The class $M_{pi}(\Gamma)$ consists of all the measures of the form ν_r, $0 \leqslant r \leqslant 1$, defined in (4.4). Moreover, let $\mu_0 = \nu_{r_0}$, $\mu_1 = \nu_{r_1} \in M_{pi}(\Gamma)$ for some $0 \leqslant r_0$, $r_1 \leqslant 1$ such that $r_0 \neq r_1$, $0 < r_0 < 1$, then

$$S_{\mu_0}(\mu_1) = r_1 \ln \left(\frac{r_1}{r_0} \right) + (1 - r_1) \ln \left(\frac{1 - r_1}{1 - r_0} \right) > 0.$$

(For the special cases $r_1 = 1$ or $r_1 = 0$, we set, as usual, $0 \ln 0 = 0$, and if $r_0 = 1$ or $r_0 = 0$, $r_0 \neq r_1$, then $S_{\mu_0}(\mu_1) = \infty$.)

Proof.

(a) This follows immediately from Theorem 2.

(b) Clearly, for each $k \geqslant 1$ we can write

$$H_{\mu_1 \| \mu_0}(\Delta_k) = E_{\mu_1} G_{\mu_1, \mu_0, k}, \tag{4.7}$$

where the measurable function $G_{\mu_1,\mu_0,k} : \Gamma \to [0,\infty]$ is given by the formula

$$G_{\mu_1,\mu_0,k}(\gamma) = \ln \frac{\mu_1\,(A_{[-Q(k),k]}(\varphi))}{\mu_0\,(A_{[-Q(k),k]}(\varphi))}$$

provided $\gamma \in A_{[-Q(k),k]}(\varphi)$, for each $\varphi \in \Gamma_{[-Q(k),k]}$, while $H_{\mu_1\|\mu_0}(\Delta_k)$ is defined in Section 2. Therefore, by (4.4),

$$\begin{aligned} G_{\mu_1,\mu_0,k}(\gamma) &= \ln \frac{\nu_{r_1}\,(A_{[-Q(k),k]}(\varphi))}{\nu_{r_0}\,(A_{[-Q(k),k]}(\varphi))} = \ln \frac{\prod_{-Q(k)\leqslant z\leqslant k} D^z_{r_1}(\gamma_z)}{\prod_{-Q(k)\leqslant z\leqslant k} D^z_{r_0}(\gamma_z)} \\ &= \sum_{1\leqslant z\leqslant k} g_z(\gamma) + \sum_{-Q(k)\leqslant z\leqslant 0} g_z(\gamma), \end{aligned} \tag{4.8}$$

where the functions $g_z : \Gamma \to [0,\infty]$ are defined by the formula $g_z(\gamma) = \ln\left(D^z_{r_1}(\gamma_z)/D^z_{r_0}(\gamma_z)\right)$ for any $\gamma \in \Gamma$, $z \in \mathbb{Z}$. Thus, by (4.7), (4.8) and the definition of $D^z_r(s)$ we have for any $k \geqslant 1$,

$$\begin{aligned} H_{\mu_1\|\mu_0}(\Lambda_k) &= \sum_{1\leqslant z\leqslant k} E_{\mu_1} g_z + \sum_{-Q(k)\leqslant z\leqslant 0} E_{\mu_1} g_z \\ &= R_n + \sum_{1\leqslant z\leqslant k} \left(\left(\frac{1}{2} + \left(r_1 - \frac{1}{2}\right) d_z\right) \ln\left(\frac{\frac{1}{2}+\left(r_1-\frac{1}{2}\right) d_z}{\frac{1}{2}+\left(r_0-\frac{1}{2}\right) d_z}\right)\right. \\ &\qquad \left. + \left(\frac{1}{2} - \left(r_1 - \frac{1}{2}\right) d_z\right) \ln\left(\frac{\frac{1}{2}-\left(r_1-\frac{1}{2}\right) d_z}{\frac{1}{2}-\left(r_0-\frac{1}{2}\right) d_z}\right)\right), \end{aligned} \tag{4.9}$$

where

$$\begin{aligned} R_k = &\sum_{-Q(k)\leqslant z\leqslant 0} \left(\left(\frac{1}{2} + \left(r_1 - \frac{1}{2}\right) d_z\right) \ln\left(\frac{\frac{1}{2}+\left(r_1-\frac{1}{2}\right) d_z}{\frac{1}{2}+\left(r_0-\frac{1}{2}\right) d_z}\right)\right. \\ &\left. + \left(\frac{1}{2} - \left(r_1 - \frac{1}{2}\right) d_z\right) \ln\left(\frac{\frac{1}{2}-\left(r_1-\frac{1}{2}\right) d_z}{\frac{1}{2}-\left(r_0-\frac{1}{2}\right) d_z}\right)\right), \end{aligned}$$

which is well defined and finite for any $0 \leqslant r_0, r_1 \leqslant 1$, since $0 < d_z < 1$ for each $z \in \mathbb{Z}$.

Recall that $d_z = \prod_{i \geqslant z} (2p_i - 1)$ for each integer $z \in \mathbb{Z}$, and, therefore, there exists a finite limit $d_{-\infty} = \lim_{z \to -\infty} d_z$. Moreover, clearly, $0 \leqslant d_{-\infty} < 1$. Therefore, there exists a finite limit

$$\lim_{k \to \infty} \frac{1}{Q(k)} R_k = \left(\frac{1}{2} + \left(r_1 - \frac{1}{2}\right) d_{-\infty}\right) \ln \left(\frac{\frac{1}{2} + \left(r_1 - \frac{1}{2}\right) d_{-\infty}}{\frac{1}{2} + \left(r_0 - \frac{1}{2}\right) d_{-\infty}}\right) + \left(\frac{1}{2} - \left(r_1 - \frac{1}{2}\right) d_{-\infty}\right) \ln \left(\frac{\frac{1}{2} - \left(r_1 - \frac{1}{2}\right) d_{-\infty}}{\frac{1}{2} - \left(r_0 - \frac{1}{2}\right) d_{-\infty}}\right), \tag{4.10}$$

which together with (3.27) yields

$$\lim_{k \to \infty} \frac{1}{k} R_k = 0. \tag{4.11}$$

Since the assumption (4.5) immediately yields $\lim_{z \to \infty} d_z = 1$, we derive by (2.2), (4.10) and (4.11), that

$$\lim_{k \to \infty} \frac{1}{k} H_{\mu_1 \| \mu_0}(\Delta_k) = r_1 \ln \left(\frac{r_1}{r_0}\right) + (1 - r_1) \ln \left(\frac{1 - r_1}{1 - r_0}\right) < \infty$$

provided $0 < r_0 < 1$. Clearly, by Jensen inequality, $S_{\mu_0}(\mu_1) > 0$, since $r_0 \neq r_1$. Finally, if $r_0 = 1$ or $r_0 = 0$, then the condition $r_0 \neq r_1$ yields

$$\lim_{k \to \infty} \frac{1}{k} H_{\mu_1 \| \mu_0}(\Delta_k) = \lim_{k \to \infty} \frac{1}{k} H_{\mu_1 \| \mu_0}(\Delta_k) = \infty. \quad \blacksquare$$

Now we can reformulate Corollary 2.2 for our case.

Corollary 4.2. Let X_t be the Markov chain defined by (4.2) such that the condition (4.5) is satisfied, and $\mu_0 \in M_{pi}(\Gamma)$. For any closed with respect to the week topology subset K of $M(\Gamma)$,

$$\limsup_{T \to \infty} \frac{\ln P_{\mu_0}\{\zeta_T \in K\}}{T} \leqslant - \inf_{\nu \in K} S_{\mu_0}(\nu). \tag{4.12}$$

Proof. This is a particular case of Theorem 2(a). $\blacksquare$

Moreover, we actually can formulate the large deviations principle for measures from $M_{pi}(\Gamma)$, proving that the lower bounds of Donsker–Varadhan type do not hold in this case.

Corollary 4.3. Let (4.5) holds true and $\mu_1 \neq \mu_0 \in M_{pi}(\Gamma)$. Then

$$\lim_{\delta\to 0} \limsup_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U_\delta(\mu_1)\}}{T} = \lim_{\delta\to 0} \liminf_{T\to\infty} \frac{\ln P_{\mu_0}\{\zeta_T \in U_\delta(\mu_1)\}}{T}$$
$$= - S_{\mu_0}(\mu_1) < 0$$

where $U_\delta(\mu)$ is the ball of radius $\delta > 0$ (with respect to some metric on $M(\Gamma)$ corresponding to the weak topology) centered at $\mu \in M(\Gamma)$.

Proof. The assertion follows from Theorem 2 and Proposition 4.1. ∎

5. PROOF OF PROPOSITION 2.1

In order to prove Proposition 2.1 we need the following result.

Proposition 5.1. Suppose that a Markov chain X_t and a measure $\mu_0 \in M(\Gamma)$ satisfy the conditions **A0** and **A1**. Then for any integer $T, k \geqslant 1$, for each $\varepsilon > 0$ and for any Λ_k-measurable function $f : \Gamma \to (0, \infty)$ such that $E_{\mu_0} f = 1$,

$$E_{\mu_0}\left(\prod_{0\leqslant j\leqslant T} f^{\frac{1}{t(k,\varepsilon)}}\left(X_j\right)\right) \leqslant M(f)\exp\left(\frac{T+1}{t(k,\varepsilon)}\ln(1+\varepsilon)\right),$$

where $M(f) = \max_{\gamma\in\Gamma} f(\gamma)$ and $t(k,\varepsilon)$ was introduced in **A1**.

Proof. We can write $T = n\,t(k,\varepsilon) - 1 + r$, where $n = [(T+1)/t(k,\varepsilon)]$, $r < t(k,\varepsilon)$. Since $M(f) \geqslant 1$ we have

$$E_{\mu_0}\left(\prod_{0\leqslant j\leqslant T} f^{1/t(k,\varepsilon)}\left(X_j\right)\right) \leqslant M(f)\, E_{\mu_0}\left(\prod_{0\leqslant j\leqslant n\,t(k,\varepsilon)-1} f^{1/t(k,\varepsilon)}\left(X_j\right)\right). \tag{5.1}$$

Next, for each $0 \leqslant m \leqslant t(k,\varepsilon) - 1$ introduce the random variable

$$D_m = \prod_{0\leqslant j\leqslant n-1} f^{1/t(k,\varepsilon)}\left(X_{j\,t(k,\varepsilon)+m}\right).$$

Then, by Jensen's inequality,

$$
\begin{aligned}
E_{\mu_0}&\left(\prod_{0\leqslant j\leqslant n\,t(k,\varepsilon)-1} f^{\,1/t(k,\varepsilon)}\left(X_j\right)\right)\\
&= E_{\mu_0}\left(\prod_{0\leqslant m\leqslant t(k,\varepsilon)-1} D_m\right)\\
&= E_{\mu_0}\left(\exp\left(\frac{1}{t(k,\varepsilon)}\sum_{0\leqslant m\leqslant t(k,\varepsilon)-1} t(k,\varepsilon)\ln D_m\right)\right)\\
&\leqslant E_{\mu_0}\left(\frac{1}{t(k,\varepsilon)}\sum_{0\leqslant m\leqslant t(k,\varepsilon)-1}\exp\left(t(k,\varepsilon)\ln D_m\right)\right)\\
&= \frac{1}{t(k,\varepsilon)}\sum_{0\leqslant m\leqslant t(k,\varepsilon)-1} E_{\mu_0} D_m^{t(k,\varepsilon)}.
\end{aligned}
\tag{5.2}
$$

But, for each $0\leqslant m\leqslant t(k,\varepsilon)-1$, combining the definition of $D_m$with the condition **A1** (in particular, using the fact that μ_0 is an invariant measure of the Markov chain X_t), we have

$$
\begin{aligned}
E_{\mu_0} D_m^{t(k,\varepsilon)} &= E_{\mu_0}\left(\prod_{0\leqslant j\leqslant n-1} f\left(X_{j\,t(k,\varepsilon)+m}\right)\right)\\
&= E_{\mu_0}\left(\prod_{0\leqslant j\leqslant n-1} f\left(X_{j\,t(k,\varepsilon)}\right)\right)\\
&\leqslant (1+\varepsilon)^n\left(E_{\mu_0} f\right)^n = (1+\varepsilon)^n.
\end{aligned}
\tag{5.3}
$$

Now (5.1)–(5.3) yield the required inequality. ▮

Next, we can complete the proof of Proposition 2.1. Let $\mu_0\in M(\Gamma)$ be a measure satisfying the conditions **A0** and **A1**. For each $\mu\in M(\Gamma)$ and each integer $k\geqslant 1$ define the Λ_k-measurable function $G_{\mu,k}:\Gamma\to\mathbb{R}\cup\{-\infty\}$ by the formula

$$G_{\mu,k}(\gamma)=\ln\frac{\mu(A)}{\mu_0(A)}$$

provided $\gamma\in A,\ A\in\Lambda_k$.

Observe that in our case we can rewrite the definition (2.1) in the form

$$H_{\mu\|\mu_0}(\Delta_k) = \mu(G_{\mu,k}). \tag{5.4}$$

Observe, that $G_{\mu,k}$ is not, in general, a continuous function (more precisely, it can take the value $-\infty$, since $\mu(A)$ can vanish for some $A \in \Lambda_k$). For this reason, we have to define a somewhat more complicated functions $G_{\mu,k,\delta} : \Gamma \to \mathbb{R}$ (for each given $\delta > 0$) by the formula

$$G_{\mu,k,\delta}(\gamma) = \ln\left(\frac{\delta\,\mu_0(A) + (1-\delta)\,\mu(A)}{\mu_0(A)}\right) \tag{5.5}$$

provided $\gamma \in A$, $A \in \Lambda_k$. Clearly, $G_{\mu,k,\delta}$ is continuous since the partition Λ_k is open. Since the logarithmic function is concave we have for each $\gamma \in \Gamma$,

$$G_{\mu,k,\delta}(\gamma) \geqslant (1-\delta) G_{\mu,k}(\gamma) \tag{5.6}$$

and, therefore, by (5.4),

$$\mu(G_{\mu,k,\delta}) \geqslant (1-\delta)\, H_{\mu\|\mu_0}(\Delta_k) \tag{5.7}$$

for each $\delta > 0$.

Observe, that if $S_{\mu_0}(\mu) = 0$ the statement of Proposition 2.1 is obvious. Consider the case

$$0 < S_{\mu_0}(\mu) < \infty.$$

In this case, by (2.2) and the assumption **A1**, for each given $\varepsilon > 0$ we can choose $k = k(\varepsilon, \mu) \geqslant 1$ large enough such that

$$\frac{1}{t(k, \frac{\varepsilon}{4})} H_{\mu\|\mu_0}(\Delta_k) > S_{\mu_0}(\mu) - \frac{\varepsilon}{4} \tag{5.8}$$

and so setting $\delta = \delta(\varepsilon, \mu) = (\varepsilon / 4 S_{\mu_0}(\mu))$ in (5.7), we obtain

$$\frac{1}{t(k, \frac{\varepsilon}{4})} \mu(G_{\mu,k,\delta}) \geqslant (1-\delta)\left(S_{\mu_0}(\mu) - \frac{\varepsilon}{4}\right) > S_{\mu_0}(\mu) - \frac{\varepsilon}{2}. \tag{5.9}$$

Now, for each $\varepsilon > 0$ we are able to define an open (with respect to the weak topology) neighborhood $U(\mu, \varepsilon)$ of $\mu \in M(\Gamma)$ by the formula

$$U(\mu, \varepsilon) = \left\{ \nu \in M(\Gamma) : \nu(G_{\mu,k,\delta}) > \mu(G_{\mu,k,\delta}) - \frac{\varepsilon}{4} \right\}. \tag{5.10}$$

(Here, and till the end of this section, $k = k(\varepsilon, \mu)$ and $\delta = \delta(\varepsilon, \mu)$.)

In order to show that the neighborhoods $U(\mu, \varepsilon)$ are appropriate for Proposition 2.1, consider Λ_k-measurable functions $f_\varepsilon : \Gamma \to (0, \infty)$ given by the formula $f_\varepsilon = \exp(G_{\mu,k,\delta})$. By (1.1) and the definition of f_e, for each integer $T > 0$,

$$\begin{aligned} \exp\left(\frac{T}{t(k, \frac{\varepsilon}{4})} \zeta_T(G_{\mu,k,\delta}) \right) &= \exp\left(\frac{1}{t(k, \frac{\varepsilon}{4})} \sum_{j=0}^{T-1} G_{\mu,k,\delta}(X_j) \right) \\ &= \prod_{0 \leqslant j \leqslant T-1} f_\varepsilon^{1/t(k, \frac{\varepsilon}{4})}(X_j). \end{aligned} \tag{5.11}$$

On the other hand, one can verify directly that $E_{\mu_0} f_\varepsilon = 1$. Since all the conditions of Proposition 5.1 are satisfied, we obtain by (5.11) that for each integer $T \geqslant 1$,

$$\begin{aligned} &E_{\mu_0}\left(\exp\left(\frac{T}{t(k, \frac{\varepsilon}{4})} \zeta_T(G_{\mu,k,\delta}) \right) \right) \\ &\quad = E_{\mu_0}\left(\prod_{0 \leqslant j \leqslant T-1} f_\varepsilon^{\frac{1}{t(k, \frac{\varepsilon}{4})}}(X_j) \right) \\ &\quad \leqslant M(f_\varepsilon) \exp\left(\frac{T+1}{t(k, \frac{\varepsilon}{4})} \ln(1 + \frac{\varepsilon}{4}) \right), \end{aligned} \tag{5.12}$$

where $M(f_\varepsilon) = \max_{\gamma \in \Gamma} f_\varepsilon(\gamma)$ and ε is replaced by $\varepsilon/4$.

Since $t(k, \varepsilon/4) \geqslant 1$, then using (5.9), (5.10), (5.12) and the Chebyshev inequality we obtain that for each integer $T \geqslant 1$,

$$\begin{aligned} &P_{\mu_0}\{ \zeta_T \in U(\mu, \varepsilon) \} \\ &\quad = P_{\mu_0}\left\{ \exp\left(\frac{T}{t(k, \frac{\varepsilon}{4})} \zeta_T(G_{\mu,k,\delta}) \right) > \exp\left(\frac{T}{t(k, \frac{\varepsilon}{4})} \left(\mu(G_{\mu,k,\delta}) - \frac{\varepsilon}{4} \right) \right) \right\} \end{aligned}$$

$$
\begin{aligned}
&\leqslant P_{\mu_0}\left\{\exp\left(\frac{T}{t(k,\,\frac{\varepsilon}{4})}\,\zeta_T(G_{\mu,k,\delta})\right) > \exp\left(T\left(S_{\mu_0}(\mu)-\frac{3\varepsilon}{4}\right)\right)\right\}\\
&\leqslant E_{\mu_0}\left(\exp\left(\frac{T}{t(k,\,\frac{\varepsilon}{4})}\zeta_T(G_{\mu,k,\delta})\right)\right)\exp\left(-T\left(S_{\mu_0}(\mu)-\frac{3\varepsilon}{4}\right)\right)\\
&\leqslant M(f_\varepsilon)\exp\left(\frac{T+1}{t(k,\varepsilon)}\,\ln(1+\frac{\varepsilon}{4})-T\left(S_{\mu_0}(\mu_1)-\frac{3\varepsilon}{4}\right)\right).
\end{aligned}
$$

Therefore,

$$
\limsup_{T\to\infty}\frac{\ln P_{\mu_0}\{\zeta_T\in U(\mu,\,\varepsilon)\}}{T}\leqslant -S_{\mu_0}(\mu_1)+\varepsilon
$$

proving the proposition for the case $0<S_{\mu_0}(\mu)<\infty$. The case $S_{\mu_0}(\mu)=\infty$ can be proved by, essentially, the same argument with obvious modifications.

6. PCA WITH DONSKER–VARADHAN'S LARGE DEVIATIONS ESTIMATES

The phenomenon described in Section 5 is not a general feature of PCA, and we will consider in this section some simple class of PCA's for which the lower bounds of the Donsker–Varadhan type hold true, that is the corresponding upper estimates are optimal. Consider, first of all, the following general setup, somewhat similar to assumptions of ref. 14. Let X_t be a Markov chain satisfying the conditions **A0**, **H1** and **H2** of Section 2 together with the following condition

H3* For any $k\geqslant 1$, $A,\,B\in\Lambda_k$, $x,\,y\in B$,

$$
P(x,\,A)=P(y,\,A),
$$

where the sequence of finite open partitions Λ_k of Γ, $k\geqslant 1$, was introduced in **H1**.

The main example. Let X_t be a Markov chain on $\Gamma=K^{\mathbb{Z}^d}$, $d\geqslant 1$ satisfying the conditions **C1** and **C2** of Section 3. Introduce the set of cubes:

$$
\Phi(k)=\{\underline{z}'\in\mathbb{Z}^d:\|\underline{z}'\|\leqslant k\},\quad k\geqslant 1, \tag{6.1}
$$

where, recall, $\|\underline{z}\|=\max_{1\leqslant i\leqslant d}|z_i|$ for each $\underline{z}=(z_1,\ldots,z_d)\in\mathbb{Z},^d$ and replace **C3** by the following condition:

C3* There exists $n_1 \geqslant 0$ such that

$$N(\underline{z}) \subseteq \Phi(k)$$

for any $\underline{z} \in \Phi(k)$, $k > n_1$.

Define the following sequence of partitions:

$$\Lambda_k = \{A_{\Phi(k+n_1)}(\varphi) : \varphi \in K^{\Phi(k+n_1)}\}, \quad k \geqslant 1 \tag{6.2}$$

using the notations introduced at the beginning of Section 3. Obviously, the conditions **C1, C2** and **C3*** imply **A0, H1, H2** and **H3***. In order to provide a more specific illustration of the condition **C3*,** we would like to indicate the following two particular cases satisfying this condition:

(a) *The direct product case*. Let $N(\underline{z}) = \underline{z}$ for each $\underline{z} \in \mathbb{Z}^d$, that is, each site could be considered separately, and, due to **C**2, the Markov chain X_t could be considered as a direct product of Markov chains with a phase space K defined at each site. Clearly, **C3*** is satisfied with $n_1 = 0$.

(b) *The "inside oriented" case*. Let $d = 1$, and there exists $n_1 \geqslant 0$ such that $N(z) = [z - n_1, z]$ for each $z \geqslant 0$, and $N(z) = [z, z + n_1]$ for each $z < 0$. In this case, the notation (6.1) takes the form $\Phi(k) = [-k, k]$, and, moreover, $N(z) \subseteq \Phi(k)$ for each $z \in \Phi(k)$, $k > n_1$. Therefore, **C3*** is satisfied. Observe, that this case is, in some sense, opposite to the case considered in Section 3 under the condition **C3.**

Next, we return to our general setup. Let X_t satisfy the assumptions **A0, H1, H2** and **H3***. Our main objective is to obtain the lower large deviations bounds for (1.1), since the upper bounds follow by general Donsker and Varadhan results presented in ref. 8. However, it turns out that in this case it is easier to begin with a study of the large deviations on the level of pairs of empirical measures, since it enables us to use standard properties of the relative entropy, combining them with some well established large deviations results. We will need the following additional notations.

Notations. Let $M(\Gamma \times \Gamma)$ be the set of all Borel probability measures defined on $\Gamma \times \Gamma$. Next, for any $T \in \mathbb{Z}^+$ we will define *the empirical pair measure* $\Psi_T : \Omega \to M(\Gamma \times \Gamma)$ by the formula

$$\Psi_T = \frac{1}{T} \sum_{t=0}^{T-1} \delta\left(X_t, X_{t+1}\right), \tag{6.3}$$

where $\delta(x, y)$ is the unit measure concentrated at a point $(x, y) \in \Gamma \times \Gamma$. For each $\nu \in M(\Gamma \times \Gamma)$ the left and right marginal measures $\nu_L, \nu_R \in M(\Gamma)$ are defined by $\nu_{\mathrm{L}}(A) = \nu(A \times \Gamma)$ and $\nu_{\mathrm{R}}(A) = \nu(\Gamma \times A)$. Next, we will introduce the set of *the measures with symmetrical marginal distributions* $M_S = \{\nu \in M(\Gamma \times \Gamma) : \nu_{\mathrm{L}} = \nu_{\mathrm{R}}\}$. Furthermore, for each $\nu \in M(\Gamma \times \Gamma)$ we define $\nu^P \in M(\Gamma \times \Gamma)$ by the formula

$$\nu^P(B \times A) = \int_B P(x, A)\, \nu_L(dx). \tag{6.4}$$

Let $D(\nu \| \nu^P)$ be *the divergence of* ν *with respect to* ν^P (see ref. 11), which is also known as the relative entropy or the Kullback–Leibler information in different applications. More precisely, we will use the following definition of the divergence: if $\nu \ll \nu^P$, then

$$D(\nu \| \nu^P) = \int_{\Gamma \times \Gamma} \rho \ln \rho \, d\nu^P = \int_{\Gamma \times \Gamma} \ln \rho \, d\nu, \tag{6.5}$$

where ρ is the Radon–Nikodym derivative of ν with respect to ν^P, and $D(\nu \| \nu^P) = \infty$, otherwise. Observe, that the assumption **H2** implies the condition $\nu \ll \nu^P$.

Next, introduce the finite Borel partition Δ_k of $\Gamma \times \Gamma$ consisting of all sets of the form $A \times B$, $A, B \in \Lambda_k$. We will consider also the divergences of ν with respect to ν^P restricted to the algebra generated by the partitions Δ_k. Namely, define

$$D_k(\nu \| \nu^P) = \int_{\Gamma \times \Gamma} \rho_k \ln \rho_k \, d\nu^P = H_{\nu \| \nu^P}(\Delta_k), \tag{6.6}$$

where ρ_k is the Radon–Nikodym derivative of ν with respect to ν^P restricted to the algebra generated by the partitions Δ_k and the relative entropy on the right hand side of (6.6) was defined by (2.1). Furthermore, applying Corollary 5.2.3 of ref. 11 together with the assumption **H1**, one has, for each $\nu \in M(\Gamma \times \Gamma)$

$$\lim_{k \to \infty} D_k(\nu \| \nu^P) = \sup_{k \geqslant 1} D_k(\nu \| \nu^P) = D(\nu \| \nu^P). \tag{6.7}$$

Remark 6.1. The definition (6.5) of the divergence is equivalent to another definition given in ref. 11, formula (5.2.10) and Lemma 5.2.3.

Remark 6.2. For a more comprehensive discussion of large deviations for the empirical pair measure, and, more generally, for the *multivariate empirical measures* and their connection to the relative entropy, see Ellis[12] and references there (observe that our notation ν^P corresponds to the notation $\mu_1 \otimes \pi$ of ref. 14 provided μ_1 is the left marginal of $\nu \in M(\Gamma \times \Gamma)$). Some facts concerning the empirical pair measure for PCA are provided in Section 2 of ref. 14.

The main result of this section is the following theorem.

Theorem 3. (a) For any open with respect to the week topology $W \subseteq M(\Gamma \times \Gamma)$ and for any $x \in \Gamma$,

$$\liminf_{T \to \infty} \frac{\ln P_x \{\Psi_T \in W\}}{T} \geqslant -\inf\{\widetilde{I}(\nu) : \nu \in W\},$$

where the functional $\widetilde{I}(\cdot)$ is defined by $\widetilde{I}(\nu) = D(\nu \,\|\, \nu^P)$ if $\nu \in M_S$, and $\widetilde{I}(\nu) = \infty$, otherwise.

(b) For any open with respect to the week topology $U \subseteq M(\Gamma)$ and for any $x \in \Gamma$,

$$\liminf_{T \to \infty} \frac{\ln P_x \{\zeta_T \in U\}}{T} \geqslant -\inf\{I(\mu) : \mu \in U\},$$

where $I(\mu)$ was defined by the formula (1.2).

In order to prepare some background for the proof of this theorem, as well as for Section 7, we will review some known results and introduce some auxiliary notations.

Auxiliary Markov chains. Due to **H3***, for each $k \geqslant 1$ and for any $A, B \in \Lambda_k$, we can define transition probabilities $P_k(B, A)$ by the formula

$$P_k(B, A) = P(x, A),$$

provided $x \in B$. For each $k \geqslant 1$ we can introduce the natural map $G_k : \Gamma \to \Lambda_k$ such that $G_k(x) = A \in \Lambda_k$, provided $x \in A$. Next, for each $k \geqslant 1$ we can define an auxiliary Markov chain $Y_t^k = G_k(X_t)$, $t \in \mathbb{Z}^+$, with the phase space Λ_k, such that

$$P\left\{Y_{t+1}^k = A \,\middle|\, Y_t^k = B\right\} = P_k(B, A)$$

for any $A, B \in \Lambda_k$, $t \in \mathbb{Z}^+$. Clearly, for each integer $m \geqslant 1$ and for each sequence of sets $A_0, A_1, \ldots, A_m \in \Lambda_k$,

$$P_x\,\{X_1 \in A_1, \ldots, X_m \in A_m\} = P_{A_0}\left\{Y_1^k = A_1, \ldots, Y_m^k = A_m\right\}$$

provided $x \in A_0$. For each Markov chain Y_t^k define the sequence of the pair empirical measures

$$\Psi_T^k = \frac{1}{T}\sum_{t=0}^{T-1} \delta\left(Y_t^k, Y_{t+1}^k\right), \quad T \in \mathbb{Z}^+, \tag{6.8}$$

where $\delta(A, B)$ is the unit measure on the finite set $\Lambda_k \times \Lambda_k$ concentrated on $(A, B) \in \Lambda_k \times \Lambda_k$. In other words, for each $g : \Lambda_k \times \Lambda_k \to \mathbb{R}$ and for each $T \in \mathbb{Z}^+$,

$$\Psi_T^k(g) = \frac{1}{T}\sum_{t=0}^{T-1} g\left(Y_t^k, Y_{t+1}^k\right). \tag{6.9}$$

Clearly, there exists a natural one-to-one correspondence between the sets Δ_k and $\Lambda_k \times \Lambda_k$.Furthermore, observe, that for each $k \geqslant 1$ and for each Δ_k- measurable function $f : \Gamma \times \Gamma \to \mathbb{R}$ there exists exactly one function $\widetilde{f} : \Lambda_k \times \Lambda_k \to \mathbb{R}$ such that $f(x, y) = \widetilde{f}(G_k(x), G_k(y))$, and, moreover, for each $T \in \mathbb{Z}^+$,

$$\Psi_T^k(\widetilde{f}) = \Psi_T(f). \tag{6.10}$$

Large deviations for the pair empirical measure of Markov chain with a finite state space. For the convenience of the reader, we will recall some well known results on large deviations of finite Markov chains (see, for instance, Chapter 3 of ref. 10 or Theorem 1.4 of ref. 12). Let Y_t be a Markov chain on a finite state space Λ, and let $\pi(i, j) > 0$ be the corresponding transition probabilities, that is, $\pi(i, j) = P\{Y_1 = j \mid Y_0 = i\}$ for each $i, j \in \Lambda$. Introduce, similarly to (6.4), the sequence of the pair empirical measures

$$\widetilde{\Psi}_T = \frac{1}{T}\sum_{t=0}^{T-1} \delta\left(Y_t, Y_{t+1}\right), \quad T \in \mathbb{Z}^+, \tag{6.11}$$

where $\delta(i, j)$ is the unit measure on the finite set $\Lambda \times \Lambda$ concentrated on $(i, j) \in \Lambda \times \Lambda$. Let $M(\Lambda \times \Lambda)$ be the family of all probability measures defined on the finite set $\Lambda \times \Lambda$. For each $\eta \in M(\Lambda \times \Lambda)$ let $\eta_{\mathrm{L}}, \eta_{\mathrm{R}} \in M(\Lambda)$ be the left and right marginal measures, respectively. Then, for any open with respect to the weak topology subset $\widetilde{W}$ of $M(\Lambda \times \Lambda)$ and for each $i \in \Lambda$,

$$\liminf_{T \to \infty} \frac{\ln P_i \{\widetilde{\Psi}_T \in \widetilde{W}\}}{T} \geqslant - \inf_{\eta \in \widetilde{W}} I_2(\eta), \tag{6.12}$$

where $I_2 : M(\Lambda \times \Lambda) \to [0, \infty]$ is defined by

$$I_2(\eta) = \sum \eta(i, j) \log \frac{\eta(i, j)}{\eta_{\mathrm{L}}(i) \pi(i, j)} \tag{6.13}$$

for each $\eta \in M(\Lambda \times \Lambda)$ with $\eta_{\mathrm{L}} = \eta_{\mathrm{R}}$, where the sum is taken over all the pairs $(i, j) \in \Lambda \times \Lambda$ such that $\eta_{\mathrm{L}}(i) \neq 0$ (as usual, we set $0 \log 0 = 0$). If, on the other hand, $\eta \in M(\Lambda \times \Lambda)$ is such that $\eta_{\mathrm{L}} \neq \eta_{\mathrm{R}}$, we set $I_2(\eta) = \infty$. Similarly, for any closed with respect to the weak topology subset $\widetilde{G}$ of $M(\Lambda \times \Lambda)$ and for each $i \in \Lambda$,

$$\limsup_{T \to \infty} \frac{\ln P_i \{\widetilde{\Psi}_T \in \widetilde{G}\}}{T} \leqslant - \inf_{\eta \in \widetilde{G}} I_2(\eta). \tag{6.14}$$

Now we are in a position to prove the main results of this section.

Proof of Theorem 3. (a) We will divide the proof of the statement (a) into two steps.

Step 1. Let $k \geqslant 1$. By H2, $P_k(B, A) > 0$ for any $A, B \in \Lambda_k$, $k \geqslant 1$, that is, the chains Y_t^k are irreducible Markov chains with finite state spaces. Introduce the notation $M_S^k = \{\eta \in M(\Lambda_k \times \Lambda_k) : \eta_{\mathrm{L}} = \eta_{\mathrm{R}}\}$. Then, by (6.12) and (6.13), for any open with respect to the weak topology subset $\widetilde{W}$ of $M(\Lambda_k)$ and for each $A \in \Lambda_k$,

$$\liminf_{T \to \infty} \frac{\ln P_A \{\Psi_T^k \in \widetilde{W}\}}{T} \geqslant - \inf_{\eta \in \widetilde{W}} I^k(\eta), \tag{6.15}$$

where $I^k(\eta)$ is defined for each $\eta \in M_S^k$ by

$$I^k(\eta) = \sum \eta(A, B) \log \frac{\eta(A, B)}{\eta_{\mathrm{L}}(A) P_k(A, B)}, \tag{6.16}$$

where the sum is taken over all the pairs $(A, B) \in \Lambda_k \times \Lambda_k$ such that $\eta_L(A) \neq 0$; and $I^k(\eta) = \infty$, otherwise.

Next, for each $\nu \in M(\Gamma \times \Gamma)$ define $\widetilde{\nu}_k \in M(\Lambda_k \times \Lambda_k)$ such that

$$\widetilde{\nu}_k\{(A, B)\} = \nu(A \times B) \tag{6.17}$$

for each $A, B \in \Lambda_k$. Let $\nu \in M_S$, then, clearly, $\widetilde{\nu}_k \in M_S^k$. Moreover, due to **H3***, (6.4) and (6.6),

$$D_k(\nu \,\|\, \nu^P) = \sum \nu(A \times B) \log \frac{\nu(A \times B)}{\nu_{\mathrm{L}}(A)\, P_k(A, B)}, \tag{6.18}$$

where the sum is taken over all sets of the form $A \times B$, $A, B \in \Lambda_k$ such that $\nu_{\mathrm{L}}(A) \neq 0$. Thus, due to (6.17); (6.16) and (6.18), it follows by a direct calculation, that for each $\nu \in M_S$,

$$I^k(\widetilde{\nu}_k) = D_k(\nu \,\|\, \nu^P). \tag{6.19}$$

Therefore, by (6.7), for each $\nu \in M_S$,

$$I^k(\widetilde{\nu}_k) \leqslant D(\nu \,\|\, \nu^P) = \widetilde{I}(\nu). \tag{6.20}$$

Step 2. Let $W \subseteq M(\Gamma \times \Gamma)$ be an open set with respect to the week topology. Suppose that $\inf\{\widetilde{I}(\nu') : \nu' \in W\} < \infty$, otherwise there is nothing to prove. For a given $\varepsilon > 0$ choose $\nu \in W$ such that

$$\widetilde{I}(\nu) < \inf\{\widetilde{I}(\nu') : \nu' \in W\} - \varepsilon. \tag{6.21}$$

For any $k \geqslant 1$, $\beta > 0$ define a neighborhood W_k^β of ν by

$$W_k^\beta = \bigcap_{B, A \in \Lambda_k} \{\nu' \in M(\Gamma \times \Gamma) : |\nu'(B \times A) - \nu(B \times A)| < \beta\}.$$

In view of Assumption **H1**, it is clear that we can choose k large enough and $\beta > 0$ small enough such that

$$W_0 := W_k^\beta \subset W. \tag{6.22}$$

Introduce the set $\widetilde{W}_0 \subset M(\Lambda_k \times \Lambda_k)$ in the following way:

$$\widetilde{W}_0 = \bigcap_{B, A \in \Lambda_k} \{\eta \in M(\Lambda_k \times \Lambda_k) : |\eta(B, A) - \nu(B \times A)| < \beta\}.$$

Substituting the indicators of sets $B \times A \in \Delta_k$ in place of f in (6.10), we see that $\Psi_T \in W_0$ if and only if $\underset{\sim}{\Psi}{}_T^k \in \widetilde{W}_0$. Moreover, obviously, due to the definition *of* $\widetilde{\nu}_k$ in (6.17), $\widetilde{\nu}_k \in \widetilde{W}_0$. Therefore, for any $x \in \Gamma$, $T \geqslant 0$, by (6.15), (6.20)–(6.22),

$$\begin{aligned}
&\liminf_{T\to\infty} \frac{\ln P_x \{\Psi_T \in W\}}{T} \\
&\quad \geqslant \liminf_{T\to\infty} \frac{\ln P_x \{\Psi_T \in W_0\}}{T} \\
&\quad = \liminf_{T\to\infty} \frac{\ln P_{A_0} \left\{\Psi_T^k \in \widetilde{W}_0\right\}}{T} \geqslant - \inf_{\eta \in \widetilde{W}_0} I^k(\eta) \\
&\quad \geqslant -I^k(\widetilde{\nu}_k) \geqslant -\widetilde{I}(\nu) \geqslant -\left(\inf\{\widetilde{I}\left(\nu'\right) : \nu' \in W\} - \varepsilon\right).
\end{aligned}$$

Since $\varepsilon > 0$ can be chosen arbitrary small, this inequality completes the prove of the statement (a).

(b) Using our notations, we will rewrite Definition (2.4) of ref. 8 in the form

$$\begin{aligned}
\overline{I}(\nu) = -\inf\{\log \int_{\Gamma\times\Gamma} f(x, y)\nu^P(dx, dy) \\
- \int_{\Gamma} \log f(x, y)\,\nu(dx, dy) : f \in \mathcal{U}_2\},
\end{aligned}$$

where $\nu \in M(\Gamma \times \Gamma)$ and $\mathcal{U}_2$ is the set of positive continuous functions on Γ. Next, according to the formula (2.21) of ref. 8, for each $\mu \in M(\Gamma)$,

$$I(\mu) = \inf_{\nu \in M_\mu} \overline{I}(\nu), \tag{6.23}$$

where $M_\mu = M_S \cap M_\mu^L$, $M_\mu^L = \{\nu \in M(\Gamma \times \Gamma) : \nu_L = \mu\}$.

On the other hand, using Lemma 2.1 of ref. 7 it follows that $\overline{I}(\nu) = D(\nu \,\|\, \nu^P)$ for each $\nu \in M(\Gamma \times \Gamma)$, and, therefore, for each $\nu \in M_S$,

$$\overline{I}(\nu) = \widetilde{I}(\nu), \tag{6.24}$$

where $\widetilde{I}(\nu)$ was defined in the statement (a). Now, taking into account that $\widetilde{I}(\nu)=\infty$ if $\nu\notin M_S$, together with (6.23) and (6.24), we have that for each $\mu\in M(\Gamma)$,

$$I(\mu)=\inf_{\nu\in M_\mu^L}\widetilde{I}(\nu),$$

where M_μ^L was defined just after (6.23). Finally, the assertion (b) follows from the assertion (a) by the contraction principle (see, for instance, ref. 10).

7. APPROXIMATE LARGE DEVIATIONS FOR PCA

In this section we will discuss a perturbative form of large deviations. Namely, instead of a single process, we will consider a family of Markov chains X_t^ε, $0<\varepsilon\leqslant\varepsilon_0$, such that all the Markov chains satisfy the condition **A0** with the same phase space Γ, assuming, in addition, the following general condition.

H4 There exists a Markov chain X_t satisfying the assumptions **A0, H1, H2** and **H3*,** and constants C_0, $C_1>0$ such that for any $k\geqslant 1$, $A\in\Lambda_k$, $x\in\Gamma$, $0<\varepsilon\leqslant\varepsilon_0$,

$$P_x\{X_1\in A\}(1-C_1\varepsilon)^{C_0k}\leqslant P_x\left\{X_1^\varepsilon\in A\right\}\leqslant P_x\{X_1\in A\}(1+C_1\varepsilon)^{C_0k}.$$

Example. Without loss of generality, assume that $K=\{1,2,\dots,m_0\}$. For each $0<\varepsilon\leqslant\varepsilon_0$ let X_t^ε be a PCA satisfying the conditions **C1, C2** and **C3** with $d=1$ (i.e., $\Gamma=K^Z$,) and with $N(z)=[z-n_0,\,z+n_0]$ for each $z\in\mathbb{Z}$ and for some integer $n_0>0$. Furthermore, assume that for each $z\in\mathbb{Z}$ the local probability kernel $P^{z,\varepsilon}:K^{N(z)}\times K\to(0,1)$ is defined for each $\eta\in K^{N(z)}$ and for each $j\in K$ by

$$P^{z,\varepsilon}(\eta,j)=(1-\varepsilon)\,p_{\eta_0 j}+\frac{\varepsilon}{2m_0}\sum_{0<|l|\leqslant m_0}p_{\eta_l j},$$

where $(p_{ij},\,1\leqslant i,\,j\leqslant m_0)$ is a given probability matrix with positive entries. Observe that by **C2**, for any given $\gamma\in\Gamma$, $\varphi\in\Gamma_{[-k,k]}$, $k>0$,

$$\begin{aligned}&P_\gamma\left\{X_1^\varepsilon\in A_{[-k,k]}(\varphi)\right\}\\&=P^\varepsilon(\gamma,\,A_{[-k,k]}(\varphi)).\\&=\prod_{z\in[-k,k]}P^{z,\varepsilon}\left(\pi_{N(z)}(\gamma),\,\varphi_z\right).\end{aligned}\tag{7.1}$$

Introduce the sequence of partitions

$$\Lambda_k = \{A_{[-k,\,k]}(\varphi) : \varphi \in K^{[-k,\,k]}\}$$

for each integer $k > 0$, where $A_{[a,\,b]}(\varphi)$ was defined in Section 3. Then the processes X_t^ε can be considered as a perturbation of the product Markov chain X_t defined on the same phase space with the transition probabilities defined by the formula

$$\begin{aligned} P_\gamma &\left\{X_1 \in A_{[-k,\,k]}(\varphi)\right\} \\ &= P(\gamma,\, A_{[-k,\,k]}(\varphi)) \\ &= \prod_{z\in[-k,\,k]} p_{z\,\varphi_z}. \end{aligned} \tag{7.2}$$

Clearly, X_t satisfies the conditions **A0, H1, H2** and **H3*** with respect to the sequence of partitions Λ_k(actually, it is the "direct product" case described in the main example of Section 6). Moreover, since $p_{i\,j} > 0$, $1 \leqslant i,\, j \leqslant m_0$ it is clear that the condition **H4** holds true due to (7.1) and (7.2). We conjecture that when ε is small enough the large deviations of the empirical measure of X_t^ε can be described precisely by the Donsker–Varadhan rate functional for X_t^ε which would improve approximate large deviations bounds derived below in this section which use the rate functional for X_t.

Let us return to the general case of Markov chains X_t^ε, satisfying the condition **H4**. Similarly to (1.1) and (6.3), for each $0 < \varepsilon \leqslant \varepsilon_0$ introduce the sequence of the occupational measures

$$\zeta_T^\varepsilon = \frac{1}{T}\sum_{t=0}^{T-1} \delta\left(X_t^\varepsilon\right), \quad T \in \mathbb{Z}^+,$$

where $\delta(x)$ is the unit measure concentrated at a point $x \in \Gamma$, and the sequence of the empirical pair measures

$$\Psi_T^\varepsilon = \frac{1}{T}\sum_{t=0}^{T-1} \delta\left(X_t,\, X_{t+1}\right),\ T \in \mathbb{Z}^+,$$

where $\delta(x\,,\,y)$ is the unit measure concentrated at a point $(x\,,\,y) \in \Gamma\times\Gamma$.

The main result of this section is the following.

Theorem 4. (a) Let G be a subset of $M(\Gamma \times \Gamma)$ closed with respect to the weak topology, such that $\inf\{\widetilde{I}(\nu) : \nu \in G\} < \infty$. Then for any $\delta > 0$ there exists $\varepsilon_1(\delta) > 0$, such that for each $0 < \varepsilon \leqslant \varepsilon_1(\delta)$ and for any $\gamma \in \Gamma$,

$$\limsup_{T\to\infty} \frac{\ln P_\gamma \left\{\Psi_T^\varepsilon \in G\right\}}{T} \leqslant -\inf\{\widetilde{I}(\nu) : \nu \in G\} + \delta,$$

where the functional $\widetilde{I}(\cdot)$ was defined for the process X_t in Theorem 3 of Section 6.

(b) Let W be a subset of $M(\Gamma \times \Gamma)$ open with respect to the weak topology. Then for any $\delta > 0$ there exists $\varepsilon_2(\delta) > 0$, such that for each $0 < \varepsilon \leqslant \varepsilon_2(\delta)$ and for any $\gamma \in \Gamma$,

$$\liminf_{T\to\infty} \frac{\ln P_\gamma \left\{\Psi_T^\varepsilon \in W\right\}}{T} \geqslant -\inf\{\widetilde{I}(\nu) : \nu \in W\} - \delta,$$

where the functional $\widetilde{I}(\cdot)$ was defined for the process X_t in Theorem 3 of Section 6.

Remark 7.1. If $\inf\{\widetilde{I}(\nu) : \nu \in G\} = \infty$, we can formulate the statement in the following way. For any $N > 0$ there exists $\varepsilon_1(N) > 0$, such that for each $0 < \varepsilon \leqslant \varepsilon_1(N)$ and for every $\gamma \in \Gamma$,

$$\limsup_{T\to\infty} \frac{\ln P_\gamma \left\{\Psi_T^\varepsilon \in G\right\}}{T} \leqslant -N.$$

The proof is, basically, the same.

Remark 7.2. Theorem 4 gives approximate large deviations for X_t^ε using the rate functional for the unperturbed process X_t itself since we cannot prove the precise large deviations estimates for X_t^ε with its own rate functional (though we conjecture that this can be done). Still, these approximate large deviations bounds are also useful provided δ is much smaller than $\inf\{\widetilde{I}(\nu) : \nu \in W\}$ and ε is sufficiently small.

In order to prove Theorem 4.1 we will need some auxiliary results. First of all, similarly to the Section 6, for each $k \geqslant 1$ we can introduce the natural map $G_k : \Gamma \to \Lambda_k$ such that $G_k(x) = A \in \Lambda_k$, provided $x \in A$. Next, for each $k \geqslant 1$, $0 < \varepsilon \leqslant \varepsilon_0$ we can define the processes $Y_t^{k,\varepsilon} = G_k(X_t^\varepsilon)$, $t \in \mathbb{Z}^+$, with the sample space $\Omega_k = \Lambda_k^{\mathbb{Z}^+}$. For any $k, n \geqslant 1$ let $\mathfrak{T}_{\mathfrak{k}}^{\mathfrak{n}}$ be the algebra of subsets of Ω_k generated by the events of the form $\left\{Y_1^{k,\varepsilon} = A_1, \ldots, Y_n^{k,\varepsilon} = A_n\right\}$, where $A_1, \ldots, A_n \in \Lambda_k$. Clearly, for any $k, n \geqslant 1$, each initial condition $\gamma \in \Gamma$, and every $0 < \varepsilon \leqslant \varepsilon_0$, the process X_t^ε induces a probability measure $P_\gamma^{\varepsilon,k}$ on Borel σ-algebra of Ω_k in the natural way. Namely, let $n \geqslant 1$, and

$$\mathcal{A} = \left\{Y_1^{k,\varepsilon} = A_1, \ldots, Y_n^{k,\varepsilon} = A_n\right\} \in \mathfrak{T}_{\mathfrak{k}}^{\mathfrak{n}}, \tag{7.3}$$

where $A_1, \ldots, A_n \in \Lambda_k$, then we can define

$$\begin{aligned} P_\gamma^{\varepsilon,k}(\mathcal{A}) :=\ & P_\gamma\left\{Y_1^{k,\varepsilon} = A_1, \ldots, Y_n^{k,\varepsilon} = A_n\right\} \\ =\ & P_\gamma\left\{X_1^\varepsilon \in A_1, \ldots, X_n^\varepsilon \in A_n\right\}. \end{aligned} \tag{7.4}$$

On the other hand, since the process X_t satisfies the condition **H3*** of Section 6, for each $k \geqslant 1$ and for any $A, B \in \Lambda_k$, we can define the transition probabilities $P_k(B, A)$ by the formula

$$P_k(B, A) = P_x\{X_1 \in A\}, \tag{7.5}$$

provided $x \in B$. Therefore, for any $k, n \geqslant 1$, and for each initial condition $A_0 \in \Lambda_k$, we can introduce a new independent of ε measure $P_{A_0}^k$ on the Borel σ-algebra of $\Omega_k = \Lambda_k^{\mathbb{Z}^+}$ in the following way:

$$P_{A_0}^k(\mathcal{A}) := \prod_{1 \leqslant m \leqslant n} P_k(A_{m-1}, A_m) \tag{7.6}$$

for each event $\mathcal{A}$ of the form (7.3). Clearly, the process $Y_t^{k,\varepsilon}$ is a Markov chain on a phase space Λ_k with respect to the family of measures $P_{A_0}^k$, $A_0 \in \Lambda_k$, defined on the sample space Ω_k. More precisely,

$$P_{A_0}^k\left\{Y_{t+1}^{k,\varepsilon} = A \,\middle|\, Y_t^{k,\varepsilon} = B\right\} = P_k(B, A)$$

for any $A,\ B\in\Lambda_k$, $t\in\mathbb{Z}^+$. Similarly to (6.8) , for each $k\geqslant 1$, $0<\varepsilon\leqslant\varepsilon_0$ we define the sequence of the pair empirical measures on the finite set $\Lambda_k\times\Lambda_k$ by

$$\Psi_T^{k,\varepsilon}=\frac{1}{T}\sum_{t=0}^{T-1}\delta\left(Y_t^{k,\varepsilon},\ Y_{t+1}^{k,\varepsilon}\right),\quad T\in\mathbb{Z}^+, \tag{7.7}$$

where $\delta(A,\ B)$ is the unit measure concentrated on $(A,B)\in\Lambda_k\times\Lambda_k$. Observe, that according to the definition of $Y_t^{k,\varepsilon}$, for any $(A,B)\in\Lambda_k\times\Lambda_k$,

$$\Psi_T^{k,\varepsilon}\{(A,B)\}=\Psi_T^{\varepsilon}(A\times B). \tag{7.8}$$

Furthermore , we can apply to this Markov chain the estimates (6.12) and (6.14) of Section 6, and to obtain the lower bound (similarly to the estimate (6.15)),

$$\liminf_{T\to\infty}\frac{\ln P_{A_0}^k\left\{\Psi_T^{k,\varepsilon}\in\widetilde{W}\right\}}{T}\geqslant-\inf_{\eta\in\widetilde{W}}I^k(\eta) \tag{7.9}$$

for any open with respect to the weak topology subset $\widetilde{W}$ of $M(\Lambda_k\times\Lambda_k)$, and the upper bound

$$\limsup_{T\to\infty}\frac{\ln P_{A_0}^k\left\{\Psi_T^{k,\varepsilon}\in\widetilde{G}\right\}}{T}\leqslant-\inf_{\eta\in\widetilde{G}}I^k(\eta) \tag{7.10}$$

for any open with respect to the weak topology subset $\widetilde{G}$ of $M(\Lambda_k\times\Lambda_k)$, where $I^k(\eta)$ was defined in (6.16). However, in order to apply these estimates to the measures $P_\gamma^{\varepsilon,k}$, we will need the following.

Proposition 7.1. For any $k,n\geqslant 1$, $0<\varepsilon\leqslant\varepsilon_0$, $A_0\in\Lambda_k$, $\gamma\in A_0$, and for each event $\mathcal{A}\in\mathfrak{T}_{\mathfrak{k}}^{\mathfrak{n}}$,

$$P_{A_0}^k(\mathcal{A})(1-C_1\varepsilon)^{C_0kn}\leqslant P_\gamma^{\varepsilon,k}(\mathcal{A})\leqslant P_{A_0}^k(\mathcal{A})(1+C_1\varepsilon)^{C_0kn}.$$

Proof. It is enough to prove the assertion for each $\mathcal{A} \in \mathfrak{T}_{\mathfrak{k}}^{\mathfrak{n}}$ of the form (7.3). By (7.4)

$$P_\gamma^{\varepsilon,k}(\mathcal{A}) = P_\gamma \left\{ X_1^\varepsilon \in A_1, \ldots, X_n^\varepsilon \in A_n \right\}.$$

Therefore, the statement can be easily proved by induction with respect to $n \geqslant 1$, applying the Markov property, then the assumption **H4** and, finally, (7.5) together with (7.6). ∎

The next result, is the main step in the proof of Theorem 4 (a).

Proposition 7.2. (a) For each $\nu \in M_S$ such that $\widetilde{I}(\nu) < \infty$, any $\gamma \in \Gamma$ and every $\delta > 0$ there exist an open neighborhood $U(\nu, \delta)$ of ν and $\varepsilon(\nu, \delta) > 0$ small enough such that for any $0 < \varepsilon \leqslant \varepsilon(\nu, \delta)$,

$$\limsup_{T \to \infty} \frac{\ln P_\gamma\{\Psi_T^\varepsilon \in U(\nu, \delta)\}}{T} \leqslant -\left(\widetilde{I}(\nu) - \frac{\delta}{2} \right).$$

(b) For each $\nu \in M_S$ such that $\widetilde{I}(\nu) = \infty$, every $\gamma \in \Gamma$ and any $N > 0$ there exists an open neighborhood $U(\nu, N)$ of ν and $\varepsilon(\nu, N) > 0$ small enough such that for any $0 < \varepsilon \leqslant \varepsilon(\nu, N)$,

$$\limsup_{T \to \infty} \frac{\ln P_\gamma\{\Psi_T^\varepsilon \in U(\nu, N)\}}{T} \leqslant -N.$$

(c) For any $\nu \in M(\Gamma \times \Gamma)$ such that $\nu \notin M_S$ there exist an open with respect to the week topology neighborhood $U(\nu)$ of ν and an integer $T(\nu)$ large enough such that

$$P_\gamma \left\{ \Psi_T^\varepsilon \in U(\nu) \right\} = 0 \tag{7.11}$$

for each $T \geqslant T(\nu)$, $\gamma \in \Gamma$, $0 < \varepsilon \leqslant \varepsilon_0$.

Proof. (a) By the definition, $\widetilde{I}(\nu) = D(\nu \,\|\, \nu^P)$. Therefore, according to (6.7) , for each $\alpha > 0$ we can fix an integer $k = k(\nu, \alpha)$ large enough such that

$$\infty > D_k(\nu \,\|\, \nu^P) > \widetilde{I}(\nu) - \alpha, \tag{7.12}$$

where $D_k(\nu \,\|\, \nu^P)$ has been defined in (6.5).

Similarly to (6.17), for the chosen $k = k(\nu, \alpha)$ define $\widetilde{\nu}_k \in M(\Lambda_k \times \Lambda_k)$ such that

$$\widetilde{\nu}_k\{(A, B)\} = \nu(A \times B)$$

for each $A, B \in \Lambda_k$, and consider an auxiliary set $\widetilde{W}_k^\beta \subset M(\Lambda_k \times \Lambda_k)$ defined by

$$\widetilde{W}_k^\beta = \bigcap_{B, A \in \Lambda_k} \{\eta \in M(\Lambda_k \times \Lambda_k) : |\eta\{(B, A)\} - \widetilde{\nu}_k\{(B, A)\}| < \beta\}.$$

Next, by (6.19),

$$I^k(\widetilde{\nu}_k) = D_k(\nu \,\|\, \nu^P). \tag{7.13}$$

Furthermore, by (6.16), the functional $I^k(\eta)$ can be considered as a continuous function of the variables $\eta\{(B, A)\}$, $A, B \in \Lambda_k$, provided $\eta \in M_S^k$ (see the proof of Theorem 3 in Section 6), while $I^k(\eta) = \infty$ for $\eta \notin M_S^k$. Therefore, by (7.13), the definition of $\widetilde{W}_k^\beta$ and by (7.12) we can choose $\beta = \beta(\nu, \alpha) > 0$ small enough, such that

$$I^k(\eta) > I^k(\widetilde{\nu}_k) - \alpha = D_k(\nu \,\|\, \nu^P) - \alpha \geqslant \widetilde{I}(\nu) - 2\alpha \tag{7.14}$$

for each $\eta \in \widetilde{W}_k^\beta$.

Next, for the chosen $k = k(\nu, \alpha)$, $\beta = \beta(\nu, \alpha)$ define a neighborhood W_k^β of ν by

$$W_k^\beta = \bigcap_{B, A \in \Lambda_k} \{\nu' \in M(\Gamma \times \Gamma) : |\nu'(B \times A) - \nu(B \times A)| < \beta\}.$$

Due to (7.8), $\Psi_T^\varepsilon \in W_k^\beta$ if and only if $\Psi_T^{k,\varepsilon} \in \widetilde{W}_k^\beta$. Therefore, by the definition of the measure $P_\gamma^{\varepsilon,k}$,

$$P_\gamma\{\Psi_T^\varepsilon \in W_k^\beta\} = P_\gamma^{\varepsilon,k}\{\Psi_T^{k,\varepsilon} \in \widetilde{W}_k^\beta\},$$

which, together with Proposition 7.1 and with the fact that $\{\Psi_T^\varepsilon \in \widetilde{W}_k^\beta\} \in T_k^n$, yields for any $T \geqslant 1$, $0 < \varepsilon \leqslant \varepsilon_0$, that

$$P_\gamma\{\Psi_T^\varepsilon \in W_k^\beta\} \leqslant P_{A_0}^k\{\Psi_T^\varepsilon \in \widetilde{W}_k^\beta\}(1 + C_1\varepsilon)^{C_0 k T}, \tag{7.15}$$

provided $\gamma \in A_0$, $A_0 \in \Lambda_k$ (since the choice of k and β *is* independent of T and ε). On the other hand, by (7.10) and (7.14),

$$\limsup_{T\to\infty} \frac{\ln P^k_{A_0}\left\{\Psi^{k,\varepsilon}_T \in \widetilde{W}^\beta_k\right\}}{T} \leqslant -\inf_{\eta\in\widetilde{W}^\beta_k} I^k(\eta) \leqslant -(\widetilde{I}(\nu)-2\alpha). \quad (7.16)$$

Now, the statement follows immediately from (7.15) and (7.16), setting $\alpha = \delta/8$ and $U(\nu,\delta) = W^\beta_k$

(b) In this case

$$\lim_{k\to\infty} D_k(\nu \,\|\, \nu^P) = \infty.$$

Therefore, in place of (7.12), we can fix an integer $k = k(\nu, M)$ large enough such that

$$D_k(\nu \,\|\, \nu^P) > M+1. \quad (7.17)$$

Now the proof proceeds just as in the previous case.

(c) See Proposition 2.1 of ref. 14. ∎

Proof of Theorem 4. (a) Let $\gamma \in \Gamma$ and denote

$$I_0 = \inf\{\widetilde{I}(\nu) : \nu \in G\}.$$

By Proposition 7.2, for each $\nu \in G$ and $\delta > 0$ we can pick up an open neighborhood $U_0(\nu,\delta)$ of ν and a number $\varepsilon_0(\nu,\delta) > 0$ small enough such that for each $0 < \varepsilon \leqslant \varepsilon_0(\nu,\delta)$ there exists $T(\nu,\varepsilon,\delta) > 0$ such that

$$P_\gamma\{\Psi^\varepsilon_T \in U_0(\nu,\delta)\} \leqslant \exp(-(I_0-\delta)\,T) \quad (7.18)$$

provided $T \geqslant T(\nu,\varepsilon,\delta)$. Since G is a compact, we can find a finite collection of measures ν_i, $1 \leqslant i \leqslant l$, such that

$$K \subset \bigcup_{i=1}^{l} U_0(\nu_i,\delta),$$

(where l depends on $\delta > 0$). Therefore, for a given $\delta > 0$ and for each $0 < \varepsilon \leqslant \varepsilon_1(\delta) = \min_{1 \leqslant i \leqslant l} \varepsilon_0(\nu_i, \delta)$,

$$P_\gamma \{\Psi_T^\varepsilon \in G\} \leqslant \sum_{i=1}^{l} P_\gamma \{\Psi_T^\varepsilon \in U_0(\nu_i, \delta)\} \leqslant l \exp(-(I_0 - \delta) T) \tag{7.19}$$

provided $T \geqslant \max_{1 \leqslant i \leqslant l} T(\nu_i, \varepsilon, \delta)$, and so,

$$\limsup_{T \to \infty} \frac{\ln P_\gamma \{\Psi_T^\varepsilon \in G\}}{T} \leqslant -(I_0 - \delta).$$

(b) The proof of the lower bound is completely similar to the proof of Theorem 3 (a) of Section 6, replacing the auxiliary processes Ψ_T^k by $\Psi_T^{k,\varepsilon}$, and applying Proposition 7.1 in the final step of the proof.

Corollary 7.3. (a) Let V be a subset of $M(\Gamma)$ closed with respect to the week topology, such that $\inf_{\nu \in V} I(\nu) < \infty$. Then for any $\delta > 0$ there exists $\varepsilon_3(\delta) > 0$, such that for each $0 < \varepsilon \leqslant \varepsilon_3(\delta)$ and for each $\gamma \in \Gamma$,

$$\limsup_{T \to \infty} \frac{\ln P_\gamma \left\{\zeta_T^\varepsilon \in V\right\}}{T} \leqslant -\inf_{\nu \in V} I(\nu) + \delta.$$

(b) Let U be a subset of $M(\Gamma)$ open with respect to the week topology. Then for any $\delta > 0$ there exists $\varepsilon_3(\delta) > 0$, such that for each $0 < \varepsilon \leqslant \varepsilon_3(\delta)$ and for each $\gamma \in \Gamma$,

$$\liminf_{T \to \infty} \frac{\ln P_\gamma \{\zeta_T \in U\}}{T} \geqslant -\inf\{I(\mu) : \mu \in U\} - \delta,$$

where $I(\mu)$ has been defined for the process X_t by the formula (1.2).

Proof. The proof is completely similar to the contraction principle (see ref. 10). ▮

ACKNOWLEDGEMENT

This research is partially supported by a grant from Israeli Science Foundation. The second author acknowledges useful discussions with J.-D. Deuschel in Berlin during 2002–2003.

REFERENCES

1. A. de Acosta, Large deviations for vector valued functionals of a Markov chain: lower bounds, *Ann. Probab*. **16**:925–960 (1988).
2. A. de Acosta, Large deviations for empirical measures of a Markov chain, *J. Theor. Probab.* **3**:395–431 (1990).
3. P. Doukhan, *Mixing: Properties and Examples* (Springer, New York, 1994).
4. D. A. Dawson, Information flow in discrete Markov systems, *J. Appl. Probab*. **10**:63–83 (1973).
5. D. A. Dawson, Information flow in one-dimensional Markov systems, *Bull. Am. Math. Soc*. **43**:382–392 (1974).
6. J. -D. Deuschel and D. W. Strook, *Large Deviations* (Academic Press, Boston, 1989).
7. M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov processes expectations for large time, I, *Commun. Pure Appl. Math*. **28**:1–47 (1975).
8. M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov processes expectations for large time, III, *Commun. Pure Appl. Math*. **29**:389–461 (1976).
9. M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov processes expectations for large time, IV, *Commun. Pure Appl. Math*. **36** (1983).
10. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Boston, Jones and Bartlett, 1993).
11. R. M. Gray, *Entropy and Information Theory* (Springer-Verlag, New York, 1990).
12. R. S. Ellis, Large deviations for empirical measure of Markov chain with an application to the multivariate empirical measure, *Ann.Probab*. **16**:1496–1508 (1988).
13. A. Eizenberg, Y. Kifer, and B. Weiss, Large deviations for $\mathbb{Z}^d$-actions, *Commun. Math. Phys*. **164**:433–454 (1994).
14. A. Eizenberg, Y. Kifer, Large deviations for probabilistic cellular automata, *J. Stat. Phys*. **108**:1255–1280 (2002).
15. H. Föllmer, Tail structure of Markov chains on infinite product spaces, *Z. Wahr. verw. Geb*. **50**:273–288 (1979).
16. Y. Kifer, Large deviations in dynamical systems and stochastic processes, *Trans. Am. Math. Soc*. **321**:505–524 (1990).
17. J. L. Lebowitz, C. Maes, and E. Speer, Statistical mechanics of probabilistic cellular automata, *J. Stat. Phys*. **59**:117–170 (1990).
18. C. Maes and S. B. Shlosman, Ergodicity of probabilistic cellular automata: a constructive criterion, *Commun. Math. Phys*. **135**:233–251 (1991).
19. O. N. Stavskaja and I. I. Pjatetskii-Shapiro, Homogeneous networks of spontaneously active elements, *Problemy Kibernet*. **20**:91–106 (1968) (in Russian).
20. L. N. Wasserstein, Markov processes over denumerable products of spaces describing large systems of automata, *Probab. Inform. Trans*. **5**:47–52 (1969).