Journal of Statistical Physics, Vol. 108, Nos. 5/6, September 2002 (© 2002)

Large Deviations for Probabilistic Cellular Automata

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

Received November 19, 2001; accepted April 23, 2002

We consider a generalized model of a probabilistic cellular automata described by a Markov chain on an infinite dimensional space and derive certain large deviations bounds for corresponding occupational measures.

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 metric phase space Γ. Consider the sequence of occupational measures

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

where $\delta(x)$ is the unit measure concentrated at a point $x \in \Gamma$. The study of large deviations for the occupational distributions of Markov chains has been initiated by Donsker and Varadhan in ref. 1 and continued by many researches (see, for instance, refs. 2–4 and references there). Upper bounds of large deviations have been established for a general class of Feller processes on compacts by Donsker and Varadhan in refs. 5 and 6 without any additional assumptions (see also Remark 2.4 later). However, lower bounds obtained so far were based on rather strong assumptions, 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. 3, or irreducibility conditions formulated

Dedicated to D. Ruelle and Ya. Sinai.

[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

in ref. 7. Under such conditions, usually, lower and upper bounds are uniform, or at least independent of the initial conditions, and, moreover, have the same rate functionals so they are optimal for the corresponding class of processes.

Unfortunately, such assumptions, usually, are not satisfied for a large class of Markov chains, arising, for instance, in statistical mechanics. Markov chains of this type, usually called now Probabilistic Cellular Automata (PCA), were introduced more than 30 years ago by Stavskaja and Pjatetskii-Shapiro[8] as a model for a neuron network and by Wasserstein[9] as a model describing a large system of automata. Later they where studied by Dawson[1, 10] and more recently these models were considered by Maes and Shlosman.[11] Their space-time evolution was investigated in Lebowitz *et al.*[12] which leads to a different type of problems and methods since one has to deal here with more restricted classes of measures which are not only time but also space shift invariant.

Usually, PCA is described as a Markov chain X_t evolving on a phase space $\Gamma = S^{\mathbb{Z}^d}$ where S is a finite (spin) set. For $\gamma \in \Gamma$, $\underline{i} \in \mathbb{Z}^d$ denote by $\gamma_{\underline{i}}$ the $\underline{i}$ coordinate of γ. For any $B \subset \mathbb{Z}^d$ denote by π_B the natural projection from $S^{\mathbb{Z}^d}$ to S^B, and by $\mathfrak{T}_B$ the *σ-algebra* of subsets of Γ generated by the coordinate function $\gamma_{\underline{i}}$, $\underline{i} \in B$. The transition probability function of X_t is called *synchronous* if

$$P(x, \{\gamma \in \Gamma : \pi_B(\gamma) = v\}) = \prod_{\underline{i} \in B} P(x, \{\gamma \in \Gamma : \gamma_{\underline{i}} = v_{\underline{i}}\}) \tag{1.2}$$

for any $x \in \Gamma$, $B \subset \mathbb{Z}^d$, $v \in S^B$ and it is called *local* if there exists $K > 0$ such that the transition function $P(\cdot, \{\gamma \in \Gamma : \gamma_{\underline{i}} = s\})$ is $\mathfrak{T}_{N(i)}$ measurable for any fixed $\underline{i} \in \mathbb{Z}^d$, $s \in S$, where

$$N(\underline{i}) = \{\underline{j} \in \mathbb{Z}^d : \|\underline{i} - \underline{j}\| \leqslant K\}$$

and $\|\underline{z}\| := \max_{1 \leqslant i \leqslant d} |z_i|$ for $\underline{z} = (z_1,..., z_d) \in \mathbb{Z}^d$.

In this paper we deal with somewhat more general Markov chains for which we derive certain large deviations bounds for occupational measures, though our lower and upper bounds come with different rate functionals. More precisely, we will obtain some lower large deviations bounds depending on the initial distribution of the Markov chain. One of the main features of the approach presented in this paper is to consider first the empirical pair distribution, and then to apply the corresponding results to the occupational measures by means of the contraction principle. Our interest in estimates depending on the initial distribution is motivated by the fact that, in some cases, the uniform lower bounds given by Donsker–Varadhan's action functional are not valid. This phenomenon was demonstrated, for instance, in Example 1 from ref. 13 for a Markov chain with two states one of which is absorbing. Taking an infinite product of such Markov chains we obtain an example in the form of a PCA (see Example 2 in Section 4) though being somewhat degenerate it is not quite satisfactory but we believe that more interesting examples of this sort can be constructed.

Namely, in general, uniform upper bounds are optimal only if we want a rate functional independent of an initial distribution but lower bound rate functionals must depend on initial distributions unless some irreducibility conditions hold true.

Let us describe the general structure of the paper. In Section 2 we will formulate our main assumptions concerning Markov chains and will introduce our basic notations while proving some preliminary results. We also formulate in Section 2 upper bounds which can be derived from the general third level upper bounds obtained by Donsker and Varadhan for Feller processes on compacts. In order to make the paper more self-contained we will provide an independent proof of the upper bounds for the empirical pair distribution in Section 5 (describing the action functional by means of Kullback–Liebler information in the framework of our special conditions). In Section 3 we will formulate and prove the main results of this paper concerning the lower bounds, and in Section 4 we show that the traditional models of Probabilistic Cellular Automata fall in our general framework.

2. THE GENERAL SET-UP AND THE DONSKER–VARADHAN ACTION FUNCTIONAL

We assume that the following conditions are satisfied

H1. 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;

H2. 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, $k \geqslant 1$);

H3. For any $k \geqslant 1$, $x \in \Gamma$, $B \in \Lambda_k$,

$$P(x, B) := P_x\,(X_1 \in B) > 0; \tag{2.1}$$

H4. For any $k \geqslant 1$, $A \in \Lambda_k$, $B \in \Lambda_{k+1}$, $x, y \in B$,

$$P(x, A) = P(y, A), \tag{2.2}$$

and so we can define $P(B, A) = P(x, A)$ for each $x \in B$.

Remark 2.1. Observe, that conditions H1–H4 imply that the process X_t satisfies the Feller property.

Remark 2.2. Condition H2 enables us to view Γ as a space of sequences consisting of integers (spins), but this representation will not lead usually to a synchronous interactions and such modification is not helpful. On the other hand, we will see in Section 4 that traditional symbolic models of PCA fit into our set up. Furthemore, notice that due to H2 each set $A \in \Lambda_k$ is both open and closed, which enables us to take advantage of the fact that its indicator is continuous.

Without loss of generality, we can consider the sample space $(\Omega, \mathfrak{T})$, where

$$\Omega = \Gamma^{\mathbb{Z}^+}, \qquad \mathfrak{T} = \mathfrak{B}^{\mathbb{Z}^+} \tag{2.3}$$

and describe the random variables X_t: $\Omega \to \Gamma$ for any $t \in \mathbb{Z}^+$ by the formula

$$X_t(\omega) = \omega_t \tag{2.4}$$

where $\omega = (\omega_0, \omega_1,..., \omega_k,...) \in \Omega$. Denote by $M_0(\Gamma)$ and $M_0(\Gamma \times \Gamma)$ the sets of the probability Borel measures defined on Γ and $\Gamma \times \Gamma$, respectively, both equiped with the weak topology (which is the only topology we consider here on spaces of measures). For any $\mu \in M_0(\Gamma \times \Gamma)$ the left and right marginal measures $\mu_L, \mu_R \in M_0(\Gamma)$ are defined by $\mu_L(A) = \mu(A \times \Gamma)$ and $\mu_R(A) = \mu(\Gamma \times A)$. Next, we will introduce the set of *the measures with symmetrical marginal distributions*

$$M_S = \{\mu \in M_0(\Gamma \times \Gamma) : \mu_L = \mu_R\}. \tag{2.5}$$

Furthermore, for any $\nu \in M_0(\Gamma)$ we define $\nu^P \in M_0(\Gamma \times \Gamma)$ by the formula

$$\nu^P(B \times A) = \int_B P(x, A)\, \nu(dx), \tag{2.6}$$

and for any $\mu \in M_0(\Gamma \times \Gamma)$ we set $\mu^P = (\mu_L)^P$.

For any $T \in \mathbb{Z}^+$ we will define *the empirical pair distribution* Ψ_T: $\Omega \to M(\Gamma \times \Gamma)$ by the formula

$$\Psi_T = \frac{1}{T} \sum_{t=0}^{T-1} \delta(X_t, X_{t+1}) \tag{2.7}$$

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

Clearly

$$(\Psi_T)_L = \zeta_T, \tag{2.8}$$

where, recall, the occupational measures $\zeta_T: \Omega \to M(\Gamma)$ are defined by (1.1).

Remark 2.3. We find it convenient to discuss large deviations mainly on the level of the empirical pair distributions Ψ_T. Observe that, due to (2.8), the corresponding large deviation bounds for the occupational measures ζ_T follow by the contraction principle (see, for instance, ref. 3). In particular, we are going to reformulate the well known Donsker–Varadhan upper bounds in accordance with this approach, but first we would like to point out that the only measures relevant to the asymptotitcs of Ψ_T are the measures with symmetrical marginal distributions, as it follows by the next simple, but important fact.

Proposition 2.1. For any $\mu \in M_0(\Gamma \times \Gamma)$ such that $\mu \notin M_S$ there exist an open with respect to the weak topology neighborhood $U(\mu)$ of μ and an integer $T(\mu)$ large enough such that

$$P_x\{\Psi_T \in U(\mu)\} = 0 \tag{2.9}$$

for each $T \geqslant T(\mu)$, $x \in \Gamma$. Moreover, for any compact $K \subset M_0(\Gamma \times \Gamma) \setminus M_S$,

$$\lim_{T \to \infty} \frac{\ln P_x\{\Psi_T \in K\}}{T} = -\infty \tag{2.10}$$

uniformly with respect to $x \in \Gamma$.

Proof. Observe that if $\mu \notin M_S$, then there exists $f \in C(\Gamma)$ such that $\mu_L(f) \neq \mu_R(f)$, where, as usual, $C(\Gamma)$ denotes the set of all real-valued continuous functions defined on Γ. On the other hand, for each $T > 0$ and for each sample path of our Markov chain,

$$|(\Psi_T)_L\,(f) - (\Psi_T)_R(f)| \leqslant \frac{2}{T} \max\nolimits_\Gamma |f|.$$

The continuation of the proof is straightforward. ∎

Now we are in a position to introduce the action functional for the empirical pair distributions Ψ_T. Namely, set

$$\tilde{I}(\mu) = \begin{cases} D(\mu \,\|\, \mu^P) & \text{for} \quad \mu \in M_S \\ \infty, & \text{otherwise.} \end{cases} \tag{2.11}$$

Here $D(\mu \,\|\, \mu^P)$ *is the divergence of* μ with respect to μ^P (see ref. 14), which is also known as the relative entropy or the Kullback–Leibler information in different applications.

Recall (see ref. 14, Lemma 5.2.3), that if $\mu \ll \mu^P$, then

$$D(\mu \,\|\, \mu^P) = \int_{\Gamma \times \Gamma} \rho \ln \rho \, d\mu^P = \int_{\Gamma \times \Gamma} \ln \rho \, d\mu \tag{2.12}$$

where ρ is the Radon–Nikodym derivative of μ with respect to μ^P; otherwise (if μ is not absolutely continuous with respect to μ^P)

$$D(\mu \,\|\, \mu^P) = \infty. \tag{2.13}$$

As it was pointed out in the Introduction, the results formulated in Theorem 1 below can be derived from Donsker and Varadhan estimates for general Feller processes on compacts (see Remark 2.4 just following Corollary 2.2).

Theorem 1. (a) $\tilde{I}\colon M_0(\Gamma \times \Gamma) \to [0, \infty]$ is a non-negative convex lower semi-continuous functional.

(b) For any closed with respect to the weak topology $K \subseteq M_0(\Gamma \times \Gamma)$,

$$\limsup_{T \to \infty} \frac{\ln P_x\{\Psi_T \in K\}}{T} \leqslant -\inf\{\tilde{I}(\mu) \colon \mu \in K\} \tag{2.14}$$

uniformly with respect to $x \in \Gamma$.

(c) Let $\mu \in M_S$. Then

$$\tilde{I}(\mu) = 0 \tag{2.15}$$

if and only if $\mu = \mu^P$. Moreover, if (2.15) holds true, then μ_L is an invariant measure for the kernel $P(x, \cdot)$.

Next, according to the contraction principle (see ref. 3), we obtain immediately corresponding upper bounds for the occupational measures ζ_T defining the action functional $I\colon M_0(\Gamma) \to [0, \infty]$ for any $\nu \in M_0(\Gamma)$ by the formula

$$I(\nu) = \min\{\tilde{I}(\mu) \colon \nu = \mu_L,\ \mu \in M_S\} \tag{2.16}$$

(if $\{\mu \in M_S : \nu = \mu_L\} = \phi$, we set $I(\nu) = \infty$).

Corollary 2.2. (a) $I: M_0(\Gamma) \to [0, \infty]$ is a convex lower semi-continuous functional.

(b) For any closed with respect to the weak topology subset K of $M(\Gamma)$

$$\limsup_{T \to \infty} \frac{\ln P_x\{\zeta_T \in K\}}{T} \leqslant -\inf_{\nu \in K} I(\nu) \tag{2.17}$$

uniformly with respect to $x \in \Gamma$.

(c) $I(\nu) = 0$ if and only if $\nu \in M_0(\Gamma)$ is an invariant measure for $P(x, \cdot)$. Moreover, let $M_I(\Gamma)$ be set of all invariant measures for the Markov kernel $P(\cdot, \cdot)$. Then for any open with respect to the weak topology neighborhood U of $M_I(\Gamma)$ there exist constants $C_1(U)$, $C_2(U)$ such that

$$P_x\{\zeta_T \notin U\} \leqslant C_1(U) \exp(-TC_2(U)) \tag{2.18}$$

for any $T \geqslant 0$, $x \in \Gamma$.

Remark 2.4. Notice that the upper bounds are proved by Donsker and Varadhan for Feller processes on certain even more general class of phase spaces than just compacts, but this fact is irrelevant for our set-up. Recall also that the action functional $\tilde{I}(\cdot)$ appears in a different form in the original works of Donsker and Varadhan (see ref. 5, p. 395, formula (2.4)). Moreover, $I(\nu)$ is also given in a different form (see ref. 5, p. 394, formula (2.1)). The equivalence of different forms of action functionals follows from general properties of the relative entropy. The results of ref. 5 are presented on the level of occupational measures only (similar to the corollary above, although in a little different form). On the other hand, since the estimates of ref. 6 are given on the third level, one can derive from there the upper estimates for the empirical pair measure given in Theorem 1 earlier by the contraction principle. Note that the estimates of ref. 6 are formulated for the continuous time Feller processes, but it is not difficult to reformulate them for the discrete time case. It turns out, however, that under the assumptions H1–H4 a simpler proof of Theorem 1 can be provided independently of the Donsker and Varadhan classical results. The authors found it reasonable to include this simplified proof in Section 5.

Until now we discussed the result connected with the upper bounds, but the following proposition is intended to prepare the ground for our main results concerning the estimates from below, although it has some value of its own, claiming that the action functional $I(\cdot)$ is finite only for measures $\nu \in M_0(\Gamma)$ having some very special properties.

Proposition 2.3. Let $\mu \in M_S$ be such that $\tilde{I}(\mu) < \infty$. Then there exist a Radon–Nikodym derivative $\rho = \frac{d\mu}{d\mu^P} \in L_1(\mu^P)$ and a Markov kernel $G: \Gamma \times \mathfrak{B} \to [0, 1]$ such that for μ_L-almost any $x \in \Gamma$ and for any $A \in \mathfrak{B}$,

$$G(x, A) = \int_A \rho(x, y)\, P(x, dy) \tag{2.19}$$

and

$$\mu = (\mu_L)^G \tag{2.20}$$

and, moreover, μ_L is an invariant measure with respect to the kernel G.

Proof. The existence of the Radon–Nikodym derivative $\rho \in L_1(\mu^P)$ follows immediately by (2.13), and moreover, by the definition of μ^P and by Fubini's Theorem, for any $A, B \in \mathfrak{B}$,

$$\mu(B \times A) = \int_{B \times A} \rho(x, y)\, \mu_P(dx \times dy) = \int_B \int_A \rho(x, y)\, P(x, dy)\, \mu_L(dx) \tag{2.21}$$

In particular, setting $A = \Gamma$, for each $B \in \mathfrak{B}$ we have

$$\mu_{\mathrm{L}}(B) = \mu(B \times \Gamma) = \int_B \int_\Gamma \rho(x, y)\, P(x, dy)\, \mu_L(dx). \tag{2.22}$$

Observe, that the last formula yields immediately that there exists a measurable set $\Gamma_0 \subset \Gamma$ such that $\mu_L(\Gamma_0) = 1$ and for any $x \in \Gamma_0$,

$$\int_\Gamma \rho(x, y)\, P(x, dy) = 1. \tag{2.23}$$

Therefore, for any $x \in \Gamma_0$ we can define a new probability measure $G(x, \cdot)$ using the formula (2.19). To complete the definition of the Markov kernel G, we define $G(x, \cdot)$ to be an arbitrary probability Borel measure on Γ for any $x \in \Gamma \setminus \Gamma_0$ (in the spirit of ref. 5, p. 401). Now we can rewrite (2.21) in the form

$$\mu(B \times A) = \int_B G(x, A)\, \mu_L(dx) \tag{2.24}$$

proving (2.20). Next, set in (2.24) $B=\Gamma$, then by (2.5) for any $A\in\mathfrak{B}$,

$$\mu_L(A)=\mu_R(A)=\mu(\Gamma\times A)=\int_\Gamma G(x,A)\,\mu_L(dx)$$

proving the fact that μ_L is invariant with respect to the kernel G. ∎

Corollary 2.4. If $I(\nu)<\infty$, then there exist a function $\rho\in L_1(\nu^P)$ and a Markov kernel $G\colon\Gamma\times\mathfrak{B}\to[0,1]$ such that ν is an invariant measure with respect to $G(\cdot,\cdot)$ and

$$G(x,A)=\int_A \rho(x,y)\,P(x,dy)$$

for ν- almost any $x\in\Gamma$, $A\in\mathfrak{B}$.

Proof. Follows immediately by the last proposition together with the formula (2.16). ∎

3. MAIN RESULTS: THE LOWER BOUNDS

Let $\nu_0\in M_0(\Gamma)$ be a given initial distribution, and let U be an open subset of $M_0(\Gamma\times\Gamma)$. The main purpose of this section is to estimate the probability $P_{\nu_0}\{\Psi_T\in U\}$ from below, or, more precisely, to obtain some estimates of form

$$\liminf_{T\to\infty}\frac{1}{T}\ln P_{\nu_0}\{\Psi_T\in U\}\geqslant -K(U,\nu_0)\qquad\text{for some}\quad K(U,\nu_0)\geqslant 0.$$

In contrast with the classical large deviation theory, our estimates depend on the initial distribution ν_0 which is not surprising since we assume no irreducibility conditions (cf. refs. 7 and 13). In accordance with this fact, we will treat the measure P_{ν_0} as *the reference measure* throughout this section.

Recall (see ref. 14, Section 2.3), that for any two measures $\mu_1,\mu_2\in M_0(\Gamma)$ 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

$$H_{\mu_1\,\|\,\mu_2}(\Delta)=\sum_{i=1}^n \mu_1(Q_i)\ln\frac{\mu_1(Q_i)}{\mu_2(Q_i)}\tag{3.1}$$

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

In addition to the Donsker–Varadhan functionals $\tilde{I}$ and I (see (2.11) and (2.16)) we consider the family of functionals $S_{\nu_0}\colon M_0(\Gamma) \to [0, \infty]$ (for any given initial distribution ν_0) defined by the formula

$$S_{\nu_0}(\nu) = \limsup_{n \to \infty} \frac{1}{n} H_{\nu \| \nu_0}(\Lambda_n) \tag{3.2}$$

for any $\nu \in M_0(\Gamma)$ (with Λ_n satisfying H2–H4). Now we can introduce the action functionals $K_{\nu_0}\colon M(\Gamma \times \Gamma) \to [0, \infty)$ by

$$K_{\nu_0}(\mu) = S_{\nu_0}(\mu_L) + \tilde{I}(\mu) \tag{3.3}$$

with $\tilde{I}$ given by (2.11).

Let $\nu \in M_0(\Gamma)$ be such that $I(\nu) < \infty$. Then, by Corollary 2.4, there exists a Markov kernel $G(x, \cdot)$ such that ν is an invariant measure with respect to $G(x, \cdot)$. Set

$$P_\nu^G = \nu \otimes G, \tag{3.4}$$

i.e., P_ν^G is the Borel probability measure on the measure space $(\Omega, \mathfrak{T})$ induced by the kernel G under the initial distribution ν. Introduce, as usual, the one-sided time-shift transformation $\theta\colon \Omega \to \Omega$ by the formula $\theta(\omega) = (\omega_1, \omega_2, \ldots, \omega_k, \ldots)$ for any $(\omega_0, \omega_1, \ldots, \omega_k, \ldots) \in \Omega$. It is known that P_ν^G is ergodic with respect to the transformation θ if and only if the initial measure ν is ergodic with respect to the kernel $G(\cdot, \cdot)$ (see ref. 15).

We can now state our main result.

Theorem 2. Let U be an open with respect to the weak topology neighborhood of $\mu \in M_S$ such that $\tilde{I}(\mu) < \infty$ and $G(\cdot, \cdot)$ be the corresponding Markov kernel defined in Proposition 2.3. If μ_L is ergodic with respect to $G(\cdot, \cdot)$, then

$$\liminf_{T \to \infty} \frac{1}{T} \ln P_{\nu_0}\{\Psi_T \in U\} \geqslant -K_{\nu_0}(\mu). \tag{3.5}$$

Proof. For the convenience of the reader we will divide the proof into four steps.

Step 1. For any $k \geqslant 1$, $\beta > 0$ define a neighborhood U_k^β of μ by the following formula

$$U_k^\beta = \bigcap_{B,\, A \in \Lambda_k} \{\mu' \in M_0 : |\mu'(B \times A) - \mu(B \times A)| < \beta\}.$$

In view of Assumption H2, it is clear that we can choose k large enough and $\beta > 0$ small enough such that

$$U_0 := U_k^\beta \subset U. \tag{3.6}$$

Next, for any $A, B \in \Lambda_k$ define the indicator $\chi_{B,A}\colon \Omega \to R$ by the formula

$$\chi_{B,A}(\omega) = \chi_B(\omega_0)\, \chi_A(\omega_1)$$

for any $\omega = (\omega_0, \omega_1, \dots, \omega_t, \dots)$. Recall (see Remark 2.2) that due to H2 the indicators $\chi_{B,A}$ are continuous and, moreover, the sets U_k^β are open with respect to the weak topology. Furthemore, for each $T \geqslant 1$ and $\omega \in \Omega$,

$$\Psi_T(B \times A) = \frac{1}{T} \sum_{t=0}^{T-1} \chi_{B,A}(\theta^t(\omega)), \tag{3.7}$$

where θ is the time-shift transformation introduced above. Set $\nu = \mu_L$, then by (3.4) and (2.20),

$$E_{P_\nu^G}\, \chi_{B,A} = P_\nu^G(X_0 \in B, X_1 \in A) = \int_B G(x, A)\, \nu(dx) = \mu(B \times A). \tag{3.8}$$

Now, by (3.7), (3.8), the definition of $U_0 = U_k^\beta$ and the ergodic theorem one has

$$\lim_{T \to \infty} P_\nu^G\{\Psi_T \in U_0\} = 1 \tag{3.9}$$

(since P_ν^G is ergodic with respect to the shift θ).

Step 2. One of the main ideas of this proof is to use (3.9) presenting it by means of the measure P_ν instead of P_ν^G. To do this, denote by $\mathfrak{T}_T$ the σ-algebra generated by the random variables $X_0, X_1, \dots, X_T$ for a given $T \geqslant 1$. Then, by (3.4) and (2.19), there exists a Radon–Nikodym derivative

$$D_T = \left.\frac{dP_\nu^G}{dP_\nu}\right|_{\mathfrak{T}_T} = \prod_{t=0}^{T-1} \rho(X_t, X_{t+1}). \tag{3.10}$$

Denote $\Omega_T = \{\omega: D_T > 0\}$ and $\Omega_\infty = \bigcap_{T \geqslant 1} \Omega_T$. Clearly

$$P_\nu^G(\Omega_\infty) = 1 \tag{3.11}$$

and, moreover, for each event $\mathfrak{A} \in \mathfrak{T}_T$

$$\int_{\Omega_\infty \cap \mathfrak{A}} D_T^{-1}(\omega)\, P_\nu^G(d\omega) = P_\nu(\Omega_\infty \cap \mathfrak{A}). \tag{3.12}$$

On the other hand, for each $\omega \in \Omega_\infty$ we can rewrite (3.10) in the form

$$D_T = \exp\left(\sum_{t=0}^{T-1} d \circ \theta^t \right) \tag{3.13}$$

where $d: \Omega \to \mathbb{R}$ is defined by the formula

$$d(\omega) = \ln \rho(\omega_0, \omega_1) \tag{3.14}$$

for each $\omega = (\omega_0, \omega_1, ..., \omega_t, ...) \in \Omega$ (if $\omega \notin \Omega_\infty$, set $d = 0$, for instance). Moreover, by Proposition 2.3, (3.11), (3.4), (2.12) and (2.11),

$$E_{P_\nu^G} d = \int_\Gamma \int_\Gamma \ln \rho(x, y)\, G(x, dy)\, \nu(dx) = E_\mu \ln \rho = D(\mu \,\|\, \mu^P) = \tilde{I}(\mu) < \infty. \tag{3.15}$$

In particular, $d \in L_1(P_\nu^G)$. Next, for a given $\delta > 0$ introduce the event

$$\mathfrak{A}_{T,\delta} = \{D_T \leqslant \exp(T(\tilde{I}(\mu) + \delta))\}. \tag{3.16}$$

Then, by (3.11), (3.13), (3.15) and the ergodic theorem,

$$\lim_{T \to \infty} P_\nu^G(\mathfrak{A}_{T,\delta}) = 1. \tag{3.17}$$

Therefore, (3.9) together with (3.17) imply that for any given $\alpha, \delta > 0$,

$$P_\nu^G(\{\Psi_T \in U_0\} \cap \mathfrak{A}_{T,\delta}) \geqslant 1 - \alpha \tag{3.18}$$

provided $T \geqslant T_1(\alpha, \delta)$.

Step 3. Observe that the statement of Theorem 2 is given in terms of the reference measure P_{ν_0}; therefore, we should find a way to perform the change of measure transformation. Since, generally speaking, P_ν is not absolutely continuous with respect to P_{ν_0}, we should restrict our consideration to some suitable σ-algebras. Namely, for a given integer $T \geqslant 1$, and for each sequence of sets $\underline{A} = (A_0, A_1, \ldots, A_T) \in \Lambda_k^{T+1}$ (with $k \geqslant 1$ chosen in (3.6)) we define, as usual, the cylindrical subset $\Omega(\underline{A}) = \{\omega \in \Omega : \omega_i \in A_i, 0 \leqslant i \leqslant T\}$. The partition $\{\Omega(\underline{A}) \colon \underline{A} \in \Lambda_k^{T+1}\}$ generates the finite field $\mathfrak{B}_k^T$. Clearly, by the definitions of Ψ_T and of U_0 one has

$$\{\Psi_T \in U_0\} \in \mathfrak{B}_k^T. \tag{3.19}$$

We will need, however, to use some more refined partition of Ω. Namely, for a given finite sequence of sets $\underline{B} \in \Lambda_{k+T} \times \Lambda_{k+T-1} \times \cdots \times \Lambda_{k+1} \times \Lambda_k$ (i.e., $\underline{B} = (B_0, B_1, \ldots, B_T)$ is such that $B_i \in \Lambda_{k+T-i}$, $0 \leqslant i \leqslant T$) set

$$\tilde{\Omega}(\underline{B}) = \{\omega \in \Omega : \omega_i \in B_i, 0 \leqslant i \leqslant T\} \tag{3.20}$$

and consider the finite field $\tilde{\mathfrak{B}}_k^T$ generated by the partition of Ω formed by the sets $\tilde{\Omega}(\underline{B})$ for all $\underline{B} \in \Lambda_{k+T} \times \Lambda_{k+T-1} \times \cdots \times \Lambda_{k+1} \times \Lambda_k$. Obviously, $\tilde{\mathfrak{B}}_k^T \supset \mathfrak{B}_k^T$, and so by (3.19),

$$\{\Psi_T \in U_0\} \in \tilde{\mathfrak{B}}_k^T. \tag{3.21}$$

On the other hand, by (H4) for any $\underline{B} \in \Lambda_{k+T} \times \Lambda_{k+T-1} \times \cdots \times \Lambda_{k+1} \times \Lambda_k$,

$$P_{\nu_0}(\Omega(\underline{B})) = \nu_0(B_{T+k}) \prod_{t=0}^{T-1} P(B_t, B_{t+1})$$

and

$$P_\nu(\Omega(\underline{B})) = \nu(B_{T+k}) \prod_{t=0}^{T-1} P(B_t, B_{t+1}),$$

and, therefore, for any $\omega \in \Omega$,

$$r_T = \left.\frac{dP_\nu}{dP_{\nu_0}}\right|_{\tilde{\mathfrak{B}}_k^T} = \left.\frac{d\nu}{d\nu_0}\right|_{\Lambda_{T+k}} \circ X_0 \tag{3.22}$$

(we write just r_T disregarding k, since k is constant throughout the present proof). Recall that one can derive some estimates concerning $\frac{d\nu}{d\nu_0}|_{\Lambda_{k+T}}$ by means of the corresponding relative entropy (see ref. 16, Proposition 4.4). Namely, for each $C > 0$,

$$\nu\left\{x \in \Gamma : \left.\frac{d\nu}{d\nu_0}\right|_{\Lambda_{k+T}} (x) \geqslant e^C\right\} \leqslant C^{-1}(H_{\nu \| \nu_0}(\Lambda_{k+T}) + \log 2). \tag{3.23}$$

For a given $\delta > 0, T \geqslant 1$ set

$$C(\delta, T) = T(S_{\nu_0}(\nu) + \delta)$$

and introduce the event

$$\Phi_{T,\delta} = \{\omega \in \Omega : r_T(\omega) < e^{C(\delta, T)}\} \in \tilde{\mathfrak{B}}_k^T. \tag{3.24}$$

Since ν is the marginal distribution of P_ν^G corresponding to the component $\omega_0 = X_0(\omega)$ for $\omega \in \Omega$, we have by (3.22), (3.23), (3.24) and (3.2),

$$P_\nu^G(\Omega \setminus \Phi_{T,\delta}) = \nu \left\{ \left. \frac{d\nu}{d\nu_0} \right|_{\Lambda_{k+T}} \geqslant \exp(T(S_{\nu_0}(\nu) + \delta)) \right\}$$

$$\leqslant \left(\frac{T+k}{T} \right) \frac{H_{\nu \parallel \nu_0}(\Lambda_{k+T}) + \log 2}{k+T} \frac{1}{\delta + S_{\nu_0}(\nu)} \leqslant 1 - \eta, \tag{3.25}$$

where $\eta = \frac{\delta}{2(S_{\nu_0}(\nu)+\delta)} > 0$ provided $T \geqslant T_2(\delta)$. Choose $\alpha = \alpha(\delta) = \frac{\eta}{2}$ in (3.18) and denote

$$\tilde{\mathfrak{A}}_{T,\delta} = \mathfrak{A}_{T,\delta} \cap \{\Psi_T \in U_0\} \cap \Phi_{T,\delta}. \tag{3.26}$$

Then by (3.18) and (3.25),

$$P_\nu^G(\tilde{\mathfrak{A}}_{T,\delta}) \geqslant \frac{\eta}{2}, \tag{3.27}$$

for any $T \geqslant T_3(\delta) = \max(T_1(\alpha, \delta), T_2(\delta))$ (recall that $\eta > 0$ is independent of T).

Step 4. To complete the proof, we put $\mathfrak{A} = \tilde{\mathfrak{A}}_{T,\delta}$ in (3.12), then by (3.26), (3.27), and (3.16) for each $T \geqslant T_3(\delta)$,

$$P_\nu(\{\Psi_T \in U_0\} \cap \Phi_{T,\delta}) \geqslant P_\nu(\tilde{\mathfrak{A}}_{T,\delta} \cap \Omega_\infty) = \int_{\Omega_\infty \cap \tilde{\mathfrak{A}}_{T,\delta}} D_T^{-1}(\omega)\, P_\nu^G(d\omega)$$

$$\geqslant \exp(-T(\tilde{I}(\mu) + \delta))\, P_\nu^G(\tilde{\mathfrak{A}}_{T,\delta})$$

$$\geqslant \frac{\eta}{2} \exp(-T(\tilde{I}(\mu)) + \delta)). \tag{3.28}$$

Now, by (3.21), (3.22), (3.24), and (3.28) we obtain for any $\delta > 0$ and $T \geqslant T_3(\delta)$,

$$\begin{aligned}
P_{\nu_0}\{\Psi_T \in U_0\} &\geqslant P_{\nu_0}(\{\Psi_T \in U_0\} \cap \Phi_{T,\delta}) \\
&\geqslant \int_{\{\Psi_T \in U_0\} \cap \Phi_{T,\delta}} e^{-C(\delta,T)} r_T(\omega)\, P_{\nu_0}(d\omega) \\
&= e^{-C(\delta,T)} P_\nu(\{\Psi_T \in U_0\} \cap \Phi_{T,\delta}) \\
&\geqslant \frac{\eta}{2} \exp(-T(S_{\nu_0}(\nu) + \tilde{I}(\mu) + 2\delta))
\end{aligned} \tag{3.29}$$

which together with (3.6) and (3.3) completes the proof of the theorem. ∎

Denote by $M_E(\Gamma \times \Gamma)$ the set of all measures $\mu \in M_S$ such that $\tilde{I}(\mu) < \infty$ and μ_L is ergodic with respect to the corresponding Markov kernel $G(\cdot, \cdot)$ introduced in Proposition 2.3. Now the following result is an immediate conclusion of Theorem 2.

Corollary 3.1. For any open with respect to the weak topology set $U \subset M_0(\Gamma \times \Gamma)$ and any initial distribution $\nu_0 \in M_0(\Gamma)$,

$$\liminf_{T \to \infty} \frac{1}{T} \ln P_{\nu_0}\{\Psi_T \in U\} \geqslant - \inf_{\nu \in U \cap M_E(\Gamma \times \Gamma)} K_{\nu_0}(\nu)$$

provided $U \cap M_E(\Gamma \times \Gamma) \neq \phi$.

It is especially interesting to estimate from below the rate of convergence to zero of $P_{\nu_0}\{\xi_T \in U(\nu)\}$ where ν_0 and ν are two different ergodic invariant measures of $P(x, \cdot)$ and $U(\nu)$ is some neighborhood of ν such that $\nu_0 \notin U(\nu)$. The following (formulated in a little more general way) result partially answers this question.

Corollary 3.2. Let ν be an ergodic invariant measure with respect to the kernel $P(\cdot, \cdot)$ satisfying the conditions H1–H4. Then for any initial distribution ν_0 and for any open with respect to the weak topology neighborhood U of ν we have

$$P_{\nu_0}\{\xi_T \in U\} \geqslant \exp(-(S_{\nu_0}(\nu) + \delta)\, T)$$

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

Proof. Let $\mu = \nu^P$, then $\tilde{I}(\mu) = I(\nu) = 0$. Now due to (2.8), we use the contraction principle. ∎

4. THE PROBABILISTIC CELLULAR AUTOMATA

The purpose of this section is to demonstrate the connection between our general assumptions and traditional PCA models considered, for instance, in refs. 1, 10, 11, and 17. Namely, let S be a finite set. Set $\Gamma = S^{\mathbb{Z}^d}$ for some $d \geqslant 1$. For each finite $\Phi \subset \mathbb{Z}^d$ let $\pi_\Phi: S^{\mathbb{Z}^d} \to S^\Phi$ be the natural projection. For any $\varphi \in S^\Phi$ set

$$A_\varphi^\Phi = \{\gamma \in \Gamma : \pi_\Phi(\gamma) = \varphi\}. \tag{4.1}$$

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

$$\rho(\gamma, \gamma') = \sum_{\underline{z} \in \mathbb{Z}^d} 2^{-\|\underline{z}\|} \tilde{\rho}(\gamma(\underline{z}), \gamma'(\underline{z})), \qquad \gamma, \gamma' \in \Gamma, \tag{4.2}$$

where $\tilde{\rho}(s_1, s_2) = 0$ if $s_1 = s_2$ and $\tilde{\rho}(s_1, s_2) = 1$ otherwise, for any $s_1, s_2 \in S$ (considering a configuration $\gamma \in \Gamma$ as a function $\gamma: \mathbb{Z}^d \to S$). 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

(A1) 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 $P^{\underline{z}}: S^{N(\underline{z})} \times S \to [0, 1]$. More precisely, for any $\eta \in S^{N(\underline{z})}$ a probability distribution $P^{\underline{z}}(\eta, \cdot)$ is defined on S. Recall that the elements of $N(\underline{z})$ are called *the neighbors* of $\underline{z} \in \mathbb{Z}^d$;

(A2) 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 $\varphi \in S^\Phi$,

$$P(x, A_\varphi^\Phi) = \prod_{\underline{z} \in \Phi} P^{\underline{z}}(\pi_{N(\underline{z})}(x), \varphi(\underline{z})); \tag{4.3}$$

(A3) There exists an integer $n_0 \geqslant 1$ such that for each $\underline{z} \in \mathbb{Z}^d$,

$$N(\underline{z}) \subseteq \{\underline{z}' \in \mathbb{Z}^d : \|\underline{z} - \underline{z}'\| \leqslant n_0\}. \tag{4.4}$$

Now introduce the set of cubes:

$$\Phi_k = \{\underline{z} \in \mathbb{Z}^d : \|\underline{z}\| \leqslant k n_0\}, \qquad k \geqslant 1 \tag{4.5}$$

and define the sequence of partitions

$$\Lambda_k = \{A_\varphi^{\Phi_k} : \varphi \in S^{\Phi_k}\}. \tag{4.6}$$

Clearly, if (A1)–(A3) hold true, then our general assumptions (H1), (H2)) and (H4) are satisfied with the sequence Λ_k, $k \geqslant 1$, introduced in (4.6). Moreover, if $P^{\underline{z}}(\eta, s) > 0$ for any $\underline{z} \in \mathbb{Z}^d$, $\eta \in S^{N(\underline{z})}$, and $s \in S$ then (H3) is also satisfied.

Example 1. To be more specific, consider the following simple example borrowed from refs. 9 and 17. Let $d=1$, $S=\{0,1\}$, that is $\Gamma = \{0,1\}^{\mathbb{Z}}$. Let $N(z) = \{z+1\}$ for any $z \in \mathbb{Z}$. In this case we can write $S^{N(z)} = S$, since $N(z)$ consists of one element. For the sake of simplicity, we will also use the notation $\pi_z = \pi_{\{z\}}$. Suppose that we are given a sequence of 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

$$P^z(s, s') = \begin{cases} p_z & \text{if} \quad s = s' \\ 1 - p_z & \text{otherwise.} \end{cases} \tag{4.7}$$

Clearly, the conditions (A1)–(A3) are satisfied in this case, and the Markov chain X_t is well defined. Let $0 \leqslant r \leqslant 1$. Introduce the product measure ν_r on Γ by the formula

$$\nu_r\{\gamma \in \Gamma : \pi_z(\gamma) = 1\} = \tfrac{1}{2} + (r - \tfrac{1}{2})\, d_z$$

for each $z \in \mathbb{Z}$ where $d_z = \prod_{i \geqslant z} (2p_i - 1)$. It is known that the measure ν_r is the invariant measure of X_t (see ref. 9 or ref. 17), and, therefore, by Corollary 2.2, $I(\nu_r) = 0$. Suppose that $d_0 = \prod_{i \geqslant 0} (2\,p_i - 1) > 0$. More precisely, choose a sequence of positive numbers β_i, $i \geqslant 0$, such that $\sum_{i=0}^{\infty} \beta_i < \infty$, and set

$$p_z = \frac{1 + e^{-\beta_{|z|}}}{2}. \tag{4.8}$$

In this case for $z \in \mathbb{Z}$, $d_z = \exp(-\sum_{i \geqslant z} \beta_{|i|})$ which yields $d_0 > 0$ and $\lim_{z \to \infty} d_z = 1$. Then, for instance, $\nu_1 \neq \nu_0$, and, moreover, one can derive by direct calculations that in this case $S_{\nu_0}(\nu_1) > 0$. Therefore, our Corollary 3.2 becomes relevant. Actually, we conjecture that in this case the upper large deviations bounds are also given by the action functionals $S_{\nu_0}(\cdot)$. The following easier but less interesting example seems to support our approach.

Example 2. Similarly to the previous example, let $d = 1$, $S = \{0, 1\}$, $\Gamma = \{0, 1\}^{\mathbb{Z}}$. Let $N(z) = \{z\}$ for any $z \in \mathbb{Z}$, which, combined with the condition (4.3), enables us to consider the Γ-valued Markov chain X_t as a direct product of infinitely many local S-valued independent Markov chains indexed by $z \in \mathbb{Z}$. As in the previous example, we can write $S^{N(z)} = S$. Suppose that we are given a sequence of numbers $0 < p_z < 1$, where $z \in \mathbb{Z}$, such that

$$a_0 = \prod_{\infty > z > -\infty} p_z > 0, \tag{4.9}$$

as, for instance, in (4.8). Define the local transition kernels P^z for any $z \in \mathbb{Z}$ and $s \in \{0, 1\}$ by

$$P^z(0, s) = \begin{cases} p_z & \text{if} \quad s = 0 \\ 1 - p_z & \text{if} \quad s = 1 \end{cases} \quad \text{and} \quad P^z(1, s) = \begin{cases} 0 & \text{if} \quad s = 0 \\ 1 & \text{if} \quad s = 1 \end{cases} \tag{4.10}$$

(which, actually, means that each local S-valued Markov chain behaves as in Example 1 of ref. 13). Let $\gamma_0, \gamma_1 \in \Gamma$ be such that

$$\gamma_0(z) = 0, \qquad \gamma_1(z) = 1$$

for each $z \in \mathbb{Z}$, and denote by ν_0, ν_1 the probability measures on Γ concentrated at the points γ_0, γ_1, respectively. Using either our formula (2.16) together with (5.2) and Proposition 5.1 formulated in the next section, or the original representation of Donsker and Varadhan (see, for instance, refs. 5, 2 or 3) one can easily verify that

$$I(\nu_0) = -\ln a_0 < \infty.$$

On the other hand, clearly, for any open with respect to the weak topology neighborhood U of ν_0 which does not include ν_1,

$$P_{\gamma_1}\{\zeta_T \in U\} = 0, \qquad \text{and so} \quad \lim_{T \to \infty} \frac{\ln P_{\gamma_1}\{\zeta_T \in U\}}{T} = -\infty.$$

Therefore, the Donsker–Varadhan action functional does not provide the correct lower estimate in this case. On the other hand, since, clearly, $S_{\nu_1}(\nu_0) = \infty$, our action functionals $S_\nu(\cdot)$ provide (though in a trivial way) the correct asymptotics of $P_{\gamma_1}\{\zeta_T \in U\}$. Still, this example is not completly satisfactory as a justification of our approach since the Donsker–Varadhan lower estimates fail here only due to certain degeneracies (for instance, the condition H3 does not hold). Nevertheless, it shows that our lower bounds work in some cases where Donsker–Varadhan's do not.

To conclude this section observe that our framework enables us to consider more general physical models where the transition probability function are not synchronous, including models where the phase space Γ is some compact subset of $\{0, 1\}^{\mathbb{Z}}$, i.e., some configurations are not allowed (hard core models).

5. A DIRECT PROOF OF THEOREM 1

In this section we will study properties of the action functionals $I(\cdot)$ and $\tilde{I}(\cdot)$, and will prove Theorem 1 formulated in Section 2. As it was already pointed out, this results could be proved by the contraction principle from the discrete version of the general third level upper large deviations bounds presented by Donsker and Varadhan in ref. 6, but we prefer to provide here a direct proof of the upper bounds for the empirical pair distributions.

Let us introduce some additional notations. For a given $k \geqslant 1$ define the partition

$$\Delta(k) = \{A \times B : A \in \Lambda_{k+1}, B \in \Lambda_k\}. \tag{5.1}$$

Clearly, each $\Delta(k)$ is a finite Borel partition of $\Gamma \times \Gamma$. Moreover, denote by $\mathfrak{B}^2$ the Borel σ-algebra of $\Gamma \times \Gamma$. Then, by Assumption H2, the family of partitions $\Delta(k)$, $k \geqslant 0$, generates $\mathfrak{B}^2$.

For each $\mu \in M_0(\Gamma \times \Gamma)$ introduce the action functional $\tilde{I}_k(\cdot)$ for a given $k \geqslant 1$ by

$$\tilde{I}_k(\mu) = H_{\mu \,\|\, \mu^P}(\Delta(k)) \tag{5.2}$$

(see the notation (3.1)), where $\mu^P \in M_0(\Gamma \times \Gamma)$ has been defined just after (2.6).

Remark 5.1. Due to Assumptions H3 and H4, we can write $\tilde{I}_k(\cdot)$ in other forms, which could be helpful for some situations. For a given $k \geqslant 0$ define the function $q_k: \Gamma \times \Gamma \to [0, \infty)$ by

$$q_k(x, y) = -\ln P(x, A) \tag{5.3}$$

provided $x \in \Gamma$, $y \in A$, where $A \in \Lambda_k$, and let

$$\Lambda_{k+1} = \{B_1,\dots, B_n\}, \qquad \Lambda_k = \{A_1,\dots, A_m\}.$$

(Clearly, the integers n, m depend on k). Then we can write for $\mu \in M_0(\Gamma \times \Gamma)$,

$$\tilde{I}_k(\mu) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mu(B_i \times A_j) \ln \mu(B_i \times A_j) - \sum_{i=1}^{n} \mu_L(B_i) \ln \mu_L(B_i) + \mu(q_k), \tag{5.4}$$

where $\mu(q_k) = \int_\Gamma q_k \, d\mu$. Moreover, denote by $\tilde{\Delta}_{k+1}$ the partition of $\Gamma \times \Gamma$ generated by sets of the form $B_i \times \Gamma$, $1 \leqslant i \leqslant n$. Then, by (5.4),

$$\tilde{I}_k(\mu) = \mu(q_k) + H_\mu(\tilde{\Delta}_{k+1}) - H_\mu(\Delta(k)) = \mu(q_k) - H_\mu(\Delta(k) \mid \tilde{\Delta}_{k+1}) \tag{5.5}$$

where $H_\mu(\Delta_1)$ is the entropy of a given partition Δ_1 for a measure μ, and $H_\mu(\Delta_1 \mid \Delta_2)$ is the conditional entropy of Δ_1 with respect to Δ_2 for a measure μ and for any two given partitions Δ_1, Δ_2.

The next proposition allows to approximate the action functional $\tilde{I}(\cdot)$ by means of the functionals $\tilde{I}_k(\cdot)$.

Proposition 5.1. For any $\mu \in M_S$,

$$\tilde{I}(\mu) = \sup_{k \geqslant 1} \tilde{I}_k(\mu) = \lim_{k \to \infty} \tilde{I}_k(\mu)$$

Proof. Let $\mathfrak{B}_0^2$ be the field generated by the family of partitions $\Delta(k)$, $k \geqslant 1$. Clearly, the σ-algebra $\mathfrak{B}^2$ is generated by the field $\mathfrak{B}_0^2$, and, therefore, for any $\mu \in M_0(\Gamma \times \Gamma)$, according to Lemma 2.2.3 of ref. 14, we have

$$D(\mu \parallel \mu^P) = \sup_{k \geqslant 1} H_{\mu \parallel \mu^P}(\Delta(k)) = \lim_{k \to \infty} H_{\mu \parallel \mu^P}(\Delta(k)) \tag{5.6}$$

which together with (5.2) and (2.11) yield the proposition. ∎

Now we will study the basic properties of the auxiliary functionals $\tilde{I}_k$.

Proposition 5.2. For each $k \geqslant 1$ the functional $\tilde{I}_k$: $M_0(\Gamma \times \Gamma) \to [0, \infty]$ is non-negative, continuous with respect to the weak topology, and convex.

Proof. It is well known, that $H_{\mu_1 \parallel \mu_2}(\Delta)$ is non-negative for any partition Δ and any two measures $\mu_1, \mu_2 \in M_0(\Gamma \times \Gamma)$ (see Theorem 2.3.2 of ref. 14, for example). Therefore, by (5.2), $\tilde{I}_k(\mu) \geqslant 0$ for any $\mu \in M_0(\Gamma \times \Gamma)$. The proof of the next two properties relies on the notations of Remark 5.1. Observe, that since the partitions Δ_k, Δ_{k+1} and $\Delta(k)$ are open, the indicators of sets $B_i \times \Gamma$, $\Gamma \times A_j$ and $B_i \times A_j$, $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$, are continuous functions, as well as the function q_k defined in (5.3). It view of formula (5.4), this fact implies that $\tilde{I}_k(\cdot)$ is continuous with respect to the weak topology.

Now we will prove that $\tilde{I}_k(\cdot)$ is convex. Let us introduce some additional notations for this purpose. First, define $F\colon [0,\infty)^m \to \mathbb{R}$ by the formula

$$F(\underline{x}) = \sum_{j=1}^{m} x_j \ln x_j - \left(\sum_{j=1}^{m} x_j\right) \ln \left(\sum_{j=1}^{m} x_j\right) \tag{5.7}$$

(where, as usual, $0 \ln 0 = 0$ and $\underline{x} = (x_1,\dots, x_m) \in [0,\infty)^m$). Next, for given $k \geqslant 1$ and $\mu \in M_0(\Gamma \times \Gamma)$ we will construct vectors $\underline{\mu_i} = (\mu_{i,1},\dots, \mu_{i,m}) \in [0,1]^m$ such that $\mu_{i,j} = \mu(B_i \times A_j)$ for $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant m$. Then, by (5.4),

$$\tilde{I}_k(\mu) = \sum_{i=1}^{n} F(\underline{\mu_i}) + \mu(q_k). \tag{5.8}$$

Since the vectors $\underline{\mu_i}$, as well, as the integrals $\mu(q_k)$, depend linearly on $\mu \in M_0(\Gamma \times \Gamma)$, it suffices to show that F is a convex function in the entire domain $[0,\infty)^m$. To do this we will prove that the corresponding Jacoby matrix is non-negative for any $\underline{x} \in \mathbb{R}^m$ such that $x_j > 0$, $1 \leqslant j \leqslant m$. Indeed, for any $\underline{v} = (v_1,\dots, v_m) \in \mathbb{R}^m$ by a direct computation we see that

$$\begin{aligned}\sum_{j,l=1}^{m} \frac{\partial^2 F}{\partial x_j\,\partial x_l} v_l v_j &= \sum_{j=1}^{m} \frac{v_j^2}{x_j} - \left(\sum_{j=1}^{m} v_j\right)^2 \left(\sum_{j=1}^{m} x_j\right)^{-1} \\ &= \left(\sum_{j=1}^{m} x_j\right)^{-1} \left(\left(\sum_{j=1}^{m} x_j\right)\left(\sum_{j=1}^{m} v_j^2 x_j^{-1}\right) - \left(\sum_{j=1}^{m} v_j\right)^2\right).\end{aligned}$$

Next, by the Cauchy inequality

$$\left(\sum_{j=1}^{m} v_j\right)^2 = \left(\sum_{j=1}^{m} x_j^{1/2} (v_j^2 x_j^{-1})^{1/2}\right)^2 \leqslant \left(\sum_{j=1}^{m} x_j\right)\left(\sum_{j=1}^{m} v_j^2 x_j^{-1}\right),$$

and so

$$\sum_{k,l=1}^{m} \frac{\partial^2 F}{\partial x_l\,\partial x_k} v_l v_k \geqslant 0$$

proving the fact that F is a convex function in the domain $x_j > 0$, $1 \leqslant j \leqslant m$. Finally, since F is a continuous function for all $\underline{x} \in [0,\infty)^m$, including the boundary points, we conclude, that F is convex in the entire domain $[0,\infty)^m$, completing the proof of the proposition. ∎

Corollary 5.3. $\tilde{I}\colon M_0(\Gamma\times\Gamma)\to[0,\infty]$ is a non-negative convex lower semi-continuous functional.

Proof. This follows immediately by the last proposition combined with Proposition 5.1. ▮

This result gives the statement (a) of Theorem 1 and our proof of the statement (b) there is based on the following simple propositions.

Proposition 5.4. For a given $k\geqslant 1$ let α_{ij} be a matrix such that $\sum_{j=1}^{m}\alpha_{ij}=1$ and $\alpha_{ij}>0$ for each $1\leqslant i\leqslant n$, $1\leqslant j\leqslant m$ with n and m defined in (5.4). Introduce $f_k\colon \Gamma\times\Gamma\to\mathbb{R}$ by the formula

$$f_k(x, y)=\ln\alpha_{ij}+q_k(x, y) \tag{5.9}$$

provided $x\in B_i$, $y\in A_j$, $B_i\times A_j\in\Delta(k)$ with q_k defined in (5.3). Then for any $T\geqslant 1$, $\gamma\in\Gamma$,

$$E_\gamma\exp(Tf_k(\Psi_T))=1. \tag{5.10}$$

Proof. By (2.7) for any $T\geqslant 2$,

$$\begin{aligned}E_\gamma\exp\,(Tf_k(\Psi_T)) &= E_\gamma\exp\left(\sum_{i=0}^{T-1} f_k(X_t, X_{t+1})\right)\\ &= E_\gamma\exp\left(\sum_{i=0}^{T-2} f_k(X_t, X_{t+1})\right)\exp(f_k(X_{T-1}, X_T))\\ &= E_\gamma\exp\left(\sum_{i=0}^{T-2} f_k(X_t, X_{t+1})\right)E_{X_{T-1}}\exp(f_k(X_{T-1}, X_T)).\end{aligned} \tag{5.11}$$

However, for each $x\in\Gamma$ we have by the definition of f_k and α_{ij} that

$$E_x\exp(f_k(x, X_1))=\sum_{j=1}^{m}P(x, A_j)\frac{\alpha_{ij}}{P(x, A_j)}=1 \tag{5.12}$$

(here $1\leqslant i\leqslant n$ is such that $x\in B_i$), which together with (5.11) implies that for any $T\geqslant 2$,

$$E_\gamma\exp(Tf_k(\Psi_T))=E_\gamma\exp((T-1)\,f_k(\Psi_{T-1}))=\cdots=1$$

and (5.10) follows. ▮

Next, we introduce few additional notations (for a fixed $k \geqslant 1$). First of all, set $p_{ij} = P(x, A_j)$ provided $x \in B_i$, which is well defined by H4. Next, for a given measure $\mu \in M_s$ set

$$\tilde{\mu}_{ij} = \frac{\mu(B_i \times A_j)}{\mu_L(B_i)} \tag{5.13}$$

if $\mu_L(B_i) \neq 0$. If $\mu_L(B_i) = 0$, we can define $\tilde{\mu}_{ij}$ arbitrarily, but it is convenient in this case to set $\tilde{\mu}_{ij} = p_{ij}$. Denote $f_\mu(x, y) = \ln(\frac{\tilde{\mu}_{ij}}{p_{ij}})$ for $x \in B_i$, $y \in A_j$, then by (5.13) and the definition of $\tilde{I}_k(\mu)$ (see (5.4)),

$$\tilde{I}_k(\mu) = \mu(f_\mu) = \sum_{1 \leqslant i \leqslant n,\, 1 \leqslant j \leqslant m} \mu(B_i \times A_j) \ln\left(\frac{\tilde{\mu}_{ij}}{p_{ij}}\right) \tag{5.14}$$

(as usual, we set $0 \ln 0 = 0$, and, therefore all parts of the last equality are well defined). Observe, that f_μ is not, in general, a continuous function (since $\tilde{\mu}_{ij}$ can vanish). For this reason, we will define for the latter use the functions $f_{\mu,\delta}\colon \Gamma \times \Gamma \to \mathbb{R}$ (for any given $\delta > 0$) by the formula

$$f_{\mu,\delta}(x, y) = \ln\left(\frac{\delta p_{ij} + (1-\delta)\, \tilde{\mu}_{ij}}{p_{ij}}\right) \tag{5.15}$$

for $x \in B_i$, $y \in A_j$. Observe, that since the logarithmic function is concave we have that for each $x, y \in \Gamma$,

$$f_{\mu,\delta}(x, y) \geqslant (1-\delta) \ln\left(\frac{\tilde{\mu}_{ij}}{p_{ij}}\right) = (1-\delta)\, f_\mu(x, y) \tag{5.16}$$

and, therefore, by (5.14),

$$\mu(f_{\mu,\delta}) \geqslant (1-\delta)\, \tilde{I}_k(\mu). \tag{5.17}$$

Next, we will need

Proposition 5.5. For each $\mu \in M_S$ such that $\tilde{I}(\mu) < \infty$ and each $\varepsilon > 0$ there exists an open neighborhood $U(\mu, \varepsilon)$ of μ such that

$$P_\gamma\{\Psi_T \in U(\mu, \varepsilon)\} \leqslant \exp(-T(\tilde{I}(\mu) - \varepsilon)) \tag{5.18}$$

for any $T \geqslant 1$ and $\gamma \in \Gamma$.

Proof. If $\tilde{I}(\mu)=0$, there is nothing to prove. Otherwise, for a given $\varepsilon>0$ we can choose, according to Proposition 5.1, an integer $k=k(\mu,\varepsilon)$ large enough such that

$$\tilde{I}_k(\mu)\geqslant\tilde{I}(\mu)-\frac{\varepsilon}{3}>0. \tag{5.19}$$

Next, take $\delta=\frac{\varepsilon}{3\tilde{I}_k(\mu)}$ in the definition of $f_{\mu,\delta}$ (see (5.15)). Then, by (5.17) and (5.19),

$$\mu(f_{\mu,\delta})\geqslant\tilde{I}_k(\mu)-\frac{\varepsilon}{3}>\tilde{I}(\mu)-\frac{2\varepsilon}{3}. \tag{5.20}$$

Since $f_{\mu,\delta}\in C(\Gamma\times\Gamma)$, define the open neighborhood $U(\mu,\varepsilon)$ of μ by

$$U(\mu,\varepsilon)=\left\{\mu'\in M_0(\Gamma\times\Gamma):\mu'(f_{\mu,\delta})>\mu(f_{\mu,\delta})-\frac{\varepsilon}{4}\right\}. \tag{5.21}$$

Observe, that if $\mu'\in U(\mu,\varepsilon)$, then by (5.20),

$$\mu'(f_{\mu,\delta})>\tilde{I}(\mu)-\varepsilon. \tag{5.22}$$

Consequently, for any $\gamma\in\Gamma$, $T\geqslant1$, by (5.22) and Chebyshev's inequality,

$$\begin{aligned}P_\gamma\{\Psi_T\in U(\mu,\varepsilon)\}&\leqslant P_\gamma\{\Psi_T(f_{\mu,\delta})>\tilde{I}(\mu)-\varepsilon\}\\&=P_\gamma\{\exp(T\,\Psi_T(f_{\mu,\delta}))\geqslant\exp(T(\tilde{I}(\mu)-\varepsilon))\}\\&\leqslant\exp(-T(\tilde{I}(\mu)-\varepsilon))\,E_\gamma\exp(T\Psi_T(f_{\mu,\delta})).\end{aligned} \tag{5.23}$$

But the function $f_{\mu,\delta}$ satisfies the conditions of Proposition 5.4 and, therefore,

$$E_\gamma\exp(T\,\Psi_T(f_{\mu,\delta}))=1,$$

which together with (5.23) prove the statement of the proposition. ∎

Observe, that if $\tilde{I}(\mu)=\infty$ then essentially the same proof shows that for any $C>0$ there exists an open neighborhood $U(\mu,C)$ of μ such that for any $\gamma\in\Gamma$ and T large enough,

$$P_\gamma\{\Psi_T\in U(\mu,C)\}\leqslant\exp(-TC). \tag{5.24}$$

In particular, if $\mu\in M_S$, the estimate (5.24) follows by Proposition 2.1.

Next, we can complete the proof of Theorem 1. Let

$$I_0 = \inf\{\tilde{I}(\mu)\colon \mu \in K\}.$$

Then, by Proposition 5.5 and the last observation, for each $\mu \in K$ and $\varepsilon > 0$ there exists an open neighborhood $U(\mu, \varepsilon)$ of μ and a number $T(\mu) > 0$ such that for any $\gamma \in \Gamma$,

$$P_\gamma\{\Psi_T \in U(\mu, \varepsilon)\} \leqslant \exp(-(I_0 - \varepsilon)\, T) \tag{5.25}$$

provided $T \geqslant T(\mu)$ with $T(\mu) = 1$ when $\mu \in M_S$. Since K is a compact we can find a finite set of measures μ_i, $1 \leqslant i \leqslant l$, such that

$$K \subset \bigcup_{i=1}^{l} U(\mu_i, \varepsilon).$$

Now, by (5.25) we have

$$P_\gamma\{\Psi_T \in K\} \leqslant \sum_{i=1}^{l} P_\gamma\{\Psi_T \in U(\mu_i, \varepsilon)\} \leqslant l \exp(-(I_0 - \varepsilon)\, T) \tag{5.26}$$

for T large enough, and therefore,

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

Since $\varepsilon > 0$ can be taken arbitrary small, the last estimate proves the statement. Finally, observe that the statement (c) of Theorem 1 follows immediately by Lemma 5.2.1 of ref. 14.

ACKNOWLEDGMENTS

This research is partially supported by a grant from Israeli Science Foundation.

REFERENCES

1. D. A. Dawson, Information flow in discrete Markov systems, *J. Appl. Probab.* **10**:63–83 (1973).
2. J.-D. Deuschel and D. W. Strook, *Large Deviations* (Academic Press, Boston, 1989).
3. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Jones and Bartlett, Boston, 1993).
4. Y. Kifer, Large deviations in dynamical systems and stochastic processes, *Trans. Amer. Math. Soc.* **321**:505–524 (1990).

5. M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov processes expectations for large time, III, *Comm. Pure Appl. Math.* **29**:389–461 (1976).
6. M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov processes expectations for large time, IV, *Comm. Pure Appl. Math.* **36**:183–212 (1983).
7. A. de Acosta, Large deviations for vector valued functionals of a Markov chain: Lower bounds, *Ann. Probab.* **16**:925–960 (1988).
8. O. N. Stavskaja and I. I. Pjatetskii-Shapiro, Homogeneous networks of spontaneously active elements, *Problemy Kibernet.* **20**:91–106 (1968) (in Russian).
9. L. N. Wasserstein, Markov processes over denumerable products of spaces describing large systems of automata, *Problems Inform. Transmission* **5**:47–52 (1969).
10. D. A. Dawson, Information flow in one-dimensional Markov systems, *Bull. Amer. Math. Soc.* **43**:382–392 (1974).
11. C. Maes and S. B. Shlosman, Ergodicity of probabilistic cellular automata: a constructive criterion, *Comm. Math. Phys.* **135**:233–251 (1991).
12. J. L. Lebowitz, C. Maes, and E. Speer, Statistical mechanics of probabilistic cellular automata, *J. Stat. Phys.* **59**:117–170 (1990).
13. A. de Acosta, Large deviations for empirical measures of a Markov chain, *J. Theoret. Probab.* **3**:395–431 (1990).
14. R. M. Gray, *Entropy and Information Theory* (Springer-Verlag, New York, 1990).
15. M. Rosenblatt, *Markov Processes. Structure and Asymptotic Behavior* (Springer-Verlag, Berlin, 1971).
16. A. Eizenberg, Y. Kifer, and B. Weiss, Large deviations for $\mathbb{Z}^d$-actions, *Commun. Math. Phys.* **164**:433–454 (1994).
17. H. Föllmer, Tail structure of Markov chains on infinite product spaces, *Z. Wahr. verw. Geb.* **50**:273–288 (1979).
18. M. D. Donsker and S. R. S. Varadhan, Asymptotic evaluation of certain Markov processes expectations for large time, I, *Comm. Pure Appl. Math.* **28**:1–47 (1975).