The Annals of Probability
1994, Vol. 22, No. 2, 713–748

LARGE DEVIATIONS FOR MARKOV CHAINS WITH RANDOM TRANSITIONS

BY TIMO SEPPÄLÄINEN

Ohio State University

This paper presents almost sure uniform large deviation principles for the empirical distributions and empirical processes of Markov chains with random transitions. The results are derived under assumptions that generalize assumptions earlier used for time-homogeneous chains. The rate functions for the skew chain are expressed in terms of the Donsker–Varadhan functional and relative entropy. The sample chain rates are different, but they have natural upper and lower bounds in terms of familiar rate functions.

1. Introduction. A natural generalization of classical time-homogeneous Markov chains is to allow the transition probabilities to be random in a stationary fashion. Such processes are called *Markov chains with random transitions* or *Markov chains in random environments.* Cogburn (1984, 1990, 1991) has successfully studied their properties in the framework of Hopf Markov chains. His papers contain numerous references to earlier related work. The last published paper of Orey (1991) deals with the ergodic theory of these processes.

Our object of study is the large deviation theory of such Markov chains. We seek results of the Donsker–Varadhan type, which take the following general form. The ingredients are a probability space $(\Omega, \mathcal{F}, P)$, a sequence $\{\xi_n : n \in \mathbb{Z}^+\}$ of random variables taking values in a Polish space $\mathcal{S}$, and a lower-semicontinuous *rate function* $I: \mathcal{S} \to [0, \infty]$. We then say that *$I$ governs the large deviations of $\{\xi_n\}$ under P* if the following inequalities hold for closed subsets F and open subsets G of $\mathcal{S}$:

$$\limsup_{n\to\infty} \frac{1}{n} \log P\{\xi_n \in F\} \le -\inf_{x\in F} I(x)$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log P\{\xi_n \in G\} \ge -\inf_{x\in G} I(x).$$

After the seminal work of Donsker and Varadhan (1975a, b, 1976, 1983), large deviation theorems for time-homogeneous Markov chains have appeared in de Acosta (1988, 1990), Bolthausen (1987), Deuschel and Stroock (1989), Ellis (1988), Ellis and Wyner (1989), Jain (1990), Orey (1986), Stroock (1984), Ney and Nummelin (1987) and Varadhan (1984), among others.

As is well known, the elegant and unified large deviation theory of i.i.d. random variables does not carry over to the Markovian case. To derive their

Received January 1992; revised November 1992.
AMS 1991 *subject classifications.* Primary 60F10; secondary 60J05.
Key words and phrases. Large deviations, Markov chains, random transitions, random environments.

results, the above-mentioned authors have had to impose various assumptions on the transition probabilities of the chain to guarantee strong enough ergodic behavior. The assumptions employed by Ellis (1988), Ellis and Wyner (1989) and Stroock (1984) generalize naturally to random environments and, as we shall see, imply that uniform large deviation principles hold for almost every realization of the environment.

Our proofs are based on classical techniques of large deviation theory. The upper bound comes from Chebyshev's inequality and exponential tightness, and the lower bound by a Shannon–McMillan type argument. The paper is organized as follows. Section 2 describes the model and our basic assumptions about it. Section 3 develops the rate functions and presents the large deviation theorems. To expedite the reader's way toward the important results, all the proofs are collected at the end. Section 4 contains the proofs for Section 2, and Section 5 the proofs of the large deviation theorems.

2. The model. Let (Ω, T) be a pair consisting of a set Ω and an invertible map T on Ω, and let $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ be a measurable space. For each $\omega \in \Omega$, suppose $P(\omega)$ is a *Markov transition kernel* on $\mathcal{X}$, that is, $P(\omega; \cdot, \cdot)$ is a function from $\mathcal{X} \times \mathcal{B}_{\mathcal{X}}$ into $[0, 1]$ satisfying the following:

1. For each $x \in \mathcal{X}$, $P(\omega; x, \cdot)$ is a probability measure on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$.
2. For each $A \in \mathcal{B}_{\mathcal{X}}$, $P(\omega; \cdot, A)$ is a $\mathcal{B}_{\mathcal{X}}$-measurable function on $\mathcal{X}$.

Pick a starting state $\omega \in \Omega$. Its successive iterates $\omega, T\omega, T^2\omega, \dots$ generate a sequence $P(\omega), P(T\omega), P(T^2\omega), \dots$ of Markov kernels. For $x \in \mathcal{X}$, define a probability P_x^ω on the product space $(\mathcal{X}^{\mathbb{Z}^+}, \mathcal{B}_{\mathcal{X}}^{\mathbb{Z}^+})$ by the following rule: Let $\{X_k : k \in \mathbb{Z}^+\}$ denote the coordinate variables on $\mathcal{X}^{\mathbb{Z}^+}$, and let $A_0, A_1, \dots, A_n$ be elements of $\mathcal{B}_{\mathcal{X}}$. Then

$$\begin{aligned} &P_x^\omega\{X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n\} \\ &\quad = \mathbf{1}_{A_0}(x) \int_{A_1} P(\omega; x, dx_1) \int_{A_2} P(T\omega; x_1, dx_2) \\ &\qquad \times \int_{A_3} \cdots \int_{A_{n-1}} P(T^{n-2}\omega; x_{n-2}, dx_{n-1}) P(T^{n-1}\omega; x_{n-1}, A_n). \end{aligned} \tag{2.1}$$

Under P_x^ω, the variables $\{X_k\}$ form a Markov chain, called the *sample chain*, with starting state x and time n transition probability $P(T^{n-1}\omega)$. In this sense the setup describes a time-inhomogeneous Markov chain on $\mathcal{X}$, run by (Ω, T).

These Markov chains are studied conveniently via the *skew Markov chain*. This is a time-homogeneous Markov chain on an augmented state space $E = \mathcal{X} \times \Omega$. For this to be meaningful, assume that Ω comes equipped with a σ-field $\mathcal{B}_\Omega$ such that both T and its inverse are measurable and that all transition probabilities $P(\omega; x, A)$ are $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_\Omega$-measurable as functions of (x, ω). For $z = (x, \omega) \in E$, $A \in \mathcal{B}_{\mathcal{X}}$ and $B \in \mathcal{B}_\Omega$, we define the skew transition kernel $\mathbb{P}$ by the rule

$$\mathbb{P}(z, A \times B) = P(\omega; x, A)\mathbf{1}_B(T\omega).$$

Let $Z_n = (X_n, \Theta_n)$, $n = 0, 1, 2, \ldots$, denote the coordinate variables on $E^{\mathbb{Z}^+}$, with $\mathcal{X}$- and Ω-components X_n and Θ_n, respectively. By Tulcea's theorem, there exists a probability measure $\mathbb{P}_{(x,\omega)}$ on $E^{\mathbb{Z}^+}$ that turns the coordinate process $\{Z_n\}$ into a time-homogeneous Markov chain with transition probabilities $\mathbb{P}(z, dz')$ and starting state (x, ω). Then, under $\mathbb{P}_{(x,\omega)}$, the distribution of $\{X_n\}$ is P_x^ω and $\Theta_n = T^n\omega$ almost surely.

Now for topological assumptions and definitions. We assume that $\mathcal{X}$ is a locally compact Polish space and take $\mathcal{B}_\mathcal{X}$ to be its Borel field. Let s be a complete metric on $\mathcal{X}$, and let r be the corresponding Prohorov metric on the space $\mathcal{M}_1(\mathcal{X})$ of Borel probability measures on $\mathcal{X}$:

$$r(\mu, \nu) = \inf\left\{\delta > 0 : \mu(A) \le \nu\left(A^\delta\right) + \delta \text{ for all } A \in \mathcal{B}_\mathcal{X}\right\},$$

where $A^\delta = \{x \in \mathcal{X}$: there is a $y \in A$ such that $s(x,y) < \delta\}$. Note that $\overline{A}^\delta = A^\delta$, for any $A \in \mathcal{B}_\mathcal{X}$, so this definition agrees with the usual one given in terms of closed sets; r is compatible with the *weak topology* of $\mathcal{M}_1(\mathcal{X})$ generated by the space $C_b(\mathcal{X})$ of bounded continuous functions on $\mathcal{X}$, and $(\mathcal{M}_1(\mathcal{X}), r)$ is a complete separable metric space.

Concerning (Ω, T), we assume that T is a homeomorphism on the Polish space Ω and take $\mathcal{B}_\Omega$ to be the Borel field.

Let $B(\mathcal{X})$ denote the Banach space of bounded Borel functions on $\mathcal{X}$, with the supremum norm. A Markov transition kernel Q on $\mathcal{X}$ is both an operator on $B(\mathcal{X})$ and a map from $\mathcal{X}$ into $\mathcal{M}_1(\mathcal{X})$. For $f \in B(\mathcal{X})$ and $x \in \mathcal{X}$, Qf is defined by

$$Qf(x) = \int f(y) Q(x, dy),$$

and $Q(x)$ is the measure in the above integral. We say Q is *Feller continuous* if $C_b(\mathcal{X})$ is invariant under Q or, equivalently, if the map $Q: \mathcal{X} \to \mathcal{M}_1(\mathcal{X})$ is continuous. Let $\mathcal{P}(\mathcal{X})$ denote the set of Feller continuous Markov transition kernels on $\mathcal{X}$. We think of $\mathcal{P}(\mathcal{X})$ as the space $C(\mathcal{X}, \mathcal{M}_1(\mathcal{X}))$ of continuous maps from $\mathcal{X}$ into $\mathcal{M}_1(\mathcal{X})$ and topologize it with the *compact-open topology*.

Fix a countable base $\{C_k\}$ for the topology of $\mathcal{X}$, consisting of relatively compact, open sets. For $P, Q \in \mathcal{P}(\mathcal{X})$, define

$$D(P, Q) = \sum_{k=1}^{\infty} 2^{-k} \sup_{x \in C_k} r\big(P(x), Q(x)\big).$$

As a metric on $C(\mathcal{X}, \mathcal{M}_1(\mathcal{X}))$, D metrizes uniform convergence on compacts. Note also that if $P(x) = \mu$ and $Q(x) = \nu$ for all x, then $D(P, Q) = r(\mu, \nu)$. This says that the D-topology of $\mathcal{M}_1(\mathcal{X})$, considered as a subspace of $\mathcal{P}(\mathcal{X})$, is precisely its original weak topology.

2.2 LEMMA. *$(\mathcal{P}(\mathcal{X}), D)$ is a complete separable metric space and D metrizes the compact-open topology of $\mathcal{P}(\mathcal{X})$.*

Henceforth we assume that *the transition kernels* $P(\omega)$ *appearing in* (2.1) *come from a continuous map* $P: \Omega \to \mathcal{P}(\mathcal{X})$. Then we have the following lemma.

2.3 LEMMA. *The skew transition* $\mathbb{P}$ *on* E *is Feller continuous.*

Before introducing further assumptions, here are some basic examples.

2.4 EXAMPLE (The canonical setting). A natural way to construct the dynamical system in the background is to let Ω be the space $\mathcal{P}(\mathcal{X})^{\mathbb{Z}}$ of sequences of Feller transitions on $\mathcal{X}$ with the product topology and T the shift map. By Lemma 2.2, this Ω is Polish. Let P_k, $k \in \mathbb{Z}$, denote the coordinate projections from Ω into $\mathcal{P}(\mathcal{X})$. Then $P(\omega) = P_0(\omega)$ gives the map $P: \Omega \to \mathcal{P}(\mathcal{X})$ that runs the sample chains. From the point of view of the sample chains, we could have formulated everything in terms of this concrete shift. However, for the time being we shall continue to talk about a general dynamical system, for this will be convenient for proving the large deviation principles.

2.5 EXAMPLE. Cogburn (1984, 1990, 1991) and Orey (1991) work in the canonical setting with a countable $\mathcal{X}$. To be concrete, take $\mathcal{X} = \mathbb{N}$ and equip $\mathbb{N}$ with the discrete topology, so that our topological assumptions are satisfied. Let $\mathcal{P}(\mathcal{X})$ be the set of all stochastic matrices over $\mathbb{N}$. The Prohorov metric r on $\mathcal{M}_1(\mathcal{X})$ is given by

$$r(\mu, \nu) = \tfrac{1}{2}\|\mu - \nu\|_{\mathrm{var}} = \tfrac{1}{2}\sum_{k=1}^{\infty} \left|\mu(k) - \nu(k)\right|,$$

and the metric D on $\mathcal{P}(\mathcal{X})$ corresponding to $C_k = \{k\}$, for $k \in \mathbb{N}$, is given by

$$D(P, Q) = \sum_{j=1}^{\infty} 2^{-j-1} \sum_{k=1}^{\infty} \left|P(j, k) - Q(j, k)\right|.$$

To introduce the randomness of the transitions, fix a T-invariant, ergodic probability π on Ω. In other words, $\pi \circ T^{-1} = \pi$, and if $A \in \mathcal{B}_\Omega$ satisfies $T^{-1}A = A$, then $\pi(A)$ equals 0 or 1.

For $k = 1, 2, 3, \ldots$ and $\omega \in \Omega$, define Feller transitions $P^k(\omega)$ by $P^1(\omega) = P(\omega)$ and, for $k > 1$,

$$P^k(\omega; x, A) = \int P^{k-1}(T\omega; y, A) P(\omega; x, dy).$$

In other words, $P^k(\omega; x, A) = P_x^\omega\{X_k \in A\}$, and $P^k(\omega)$ is the k-step transition of the sample chain on $\mathcal{X}$, given that the dynamical system is at ω. Now we come to a basic assumption concerning the setup.

ASSUMPTION (A). There exist a positive integer b, a T-invariant Borel subset Ω_A of Ω and a measurable function $M: \Omega \to [1, \infty)$, such that $\pi(\Omega_A) = 1$,

$\log M \in L^1(\pi)$, and the following inequality holds for all $\omega \in \Omega_A$, $x,y \in \mathcal{X}$ and $A \in \mathcal{B}_{\mathcal{X}}$:

$$P^b(\omega; x, A) \le M(\omega) P^b(\omega; y, A). \tag{2.6}$$

REMARKS.

(i) If (2.6) holds for a particular ω and b, it continues to hold for this ω if b is increased.

(ii) The T-invariance of Ω_A is included merely for convenience. It is not a real restriction, since any Borel set of full measure contains an invariant set of full measure.

(iii) In terms of the skew process, (2.6) reads, for all $\omega \in \Omega_A$, $x,y \in \mathcal{X}$ and $C \in \mathcal{B}_E$,

$$\mathbb{P}^b\big((x,\omega), C\big) \le M(\omega)\mathbb{P}^b\big((y,\omega), C\big). \tag{2.7}$$

2.8 THEOREM. *Assume* (A). *Then the following hold*:

(i) *Among the* $\mathbb{P}$-*invariant probabilities on* E, *there is a unique one with* Ω-*marginal* π. *Call it* Φ.

(ii) *The probability*

$$\mathbb{P}_\Phi = \int \mathbb{P}_{(x,\omega)} \Phi(dx, d\omega)$$

on $E^{\mathbb{Z}^+}$ *is ergodic.*

(iii) *Let* $f \in B(E)$. *Then there exists* $\Omega_f \in \mathcal{B}_\Omega$ *such that* $\pi(\Omega_f) = 1$ *and, for all* $\omega \in \Omega_f$ *and* $x \in \mathcal{X}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(Z_k) = \int f\, d\Phi$$

holds $\mathbb{P}_{(x,\omega)}$-*almost surely.*

2.9 EXAMPLE. A sequence of independent random variables with distributions generated by an ergodic process is an obvious special case of our setup and satisfies Assumption (A) trivially. Large deviation principles for such processes have appeared in Comets (1989), Baxter, Jain and Seppäläinen (1993) and Seppäläinen (1991, 1993a, b).

3. Large deviation theorems. This section studies large deviations from the ergodic behavior of Theorem 2.8(iii). We shall look at both position and process level for both the skew chain and the sample chain. Throughout, we shall use the abbreviation $I(A) = \inf_{x\in A} I(x)$ for any function I and any subset A of its domain.

3.1. *Skew chain.* At position level, the random variables of interest are the $\mathcal{M}_1(E)$-valued *empirical distributions*

$$\mathbf{L}_n = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{Z_k}.$$

To have notation for process level, let $\vec{X} = (X_0, X_1, X_3, \ldots)$ and $\vec{X}_k = \vec{X}\circ S^k = (X_k, X_{k+1}, X_{k+2}, \ldots)$, where S is the shift map on $\mathcal{X}^{\mathbb{Z}^+}$. Since the Ω-component moves deterministically, the pair $(\vec{X}_k, \Theta_k)$ contains all the information about the future of the skew process, and we think of it as a random variable with values in the space $E_\infty = \mathcal{X}^{\mathbb{Z}^+}\times\Omega$. The $\mathcal{M}_1(E_\infty)$-valued *empirical processes* $\mathbf{M}_n$ are defined by

$$\mathbf{M}_n = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{(\vec{X}_k,\Theta_k)}.$$

Let $V_b(E)$ denote the space of bounded continuous functions that map E into $[1,\infty)$. The *Donsker–Varadhan functional* $\mathbb{J}\colon\mathcal{M}_1(E)\to[0,\infty]$ for the skew transition is defined by

$$\mathbb{J}(\Gamma) = \sup\left\{\int\log\frac{u}{\mathbb{P}u}\,d\Gamma\colon u\in V_b(E)\right\}. \tag{3.1}$$

Let $\mathcal{M}_\pi(E)$ be the set of probabilities on E whose Ω-marginal is π. Define $I_E\colon\mathcal{M}_1(E)\to[0,\infty]$ by

$$I_E(\Gamma) = \begin{cases}\mathbb{J}(\Gamma), & \text{if } \Gamma\in\mathcal{M}_\pi(E),\\ \infty, & \text{otherwise.}\end{cases} \tag{3.2}$$

3.3 THEOREM. *The functional I_E is lower semicontinuous and convex. Assume* (A). *Then $I_E(\Gamma) = 0$ if and only if $\Gamma = \Phi$. Moreover, I_E has compact level sets, meaning that*

$$\{\Gamma\in\mathcal{M}_1(E)\colon I_E(\Gamma)\le l\}$$

is compact in the weak topology of $\mathcal{M}_1(E)$ for all real l.

The functional I_E governs the large deviations of $\mathbf{L}_n$ under $\mathbb{P}_{(x,\omega)}$, uniformly over $\mathcal{X}$ and π-almost surely, in the following sense: There is a Borel subset Ω_E of Ω such that $\pi(\Omega_E) = 1$ and, for all $\omega\in\Omega_E$, these large deviation bounds hold: If $F\subset\mathcal{M}_1(E)$ is closed, then

$$\limsup_{n\to\infty}\frac{1}{n}\log\sup_{x\in\mathcal{X}}\mathbb{P}_{(x,\omega)}\{\mathbf{L}_n\in F\}\le -I_E(F),$$

and if $G\subset\mathcal{M}_1(E)$ is open, then

$$\liminf_{n\to\infty}\frac{1}{n}\log\inf_{x\in\mathcal{X}}\mathbb{P}_{(x,\omega)}\{\mathbf{L}_n\in G\}\ge -I_E(G).$$

Intermediate steps between position and process levels are furnished by *multivariate* position level results. Let $d \geq 2$ be a fixed integer. The appropriate state space is now $E_d = \mathcal{X}^d \times \Omega$, and the $\mathcal{M}_1(E_d)$-valued d-variate empirical distribution is

$$\mathbf{M}_n^{(d)} = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{((X_k,\ldots,X_{k+d-1}),\Theta_k)}.$$

Denote elements of $\mathcal{X}^d$ by $\mathbf{x} = (x_0,\ldots,x_{d-1})$. For a probability measure Γ on E_d, define the probability measure $\Gamma\mathbb{P}^{(d)}$ on E_d by

$$\Gamma\mathbb{P}^{(d)}(A) = \iint \mathbf{1}_A\big((x_1,\ldots,x_{d-1},y),T\omega\big)P\big(T^{d-1}\omega;\, x_{d-1},dy\big)\Gamma(d\mathbf{x},d\omega).$$

The d-variate rate will be expressed in terms of *relative entropy*. Given two probabilities μ and ν on a measurable space, the entropy of μ relative to ν is defined by

$$H(\mu \mid \nu) = \begin{cases} \int \log f\, d\mu, & \text{if } d\mu = f\, d\nu, \\ \infty, & \text{if } \mu \ll \nu \text{ fails.} \end{cases}$$

If we want to consider the restrictions of μ and ν to some sub-σ-field $\mathcal{D}$, we write $H_{\mathcal{D}}(\mu \mid \nu)$.

3.4 Definition. Suppose $d \geq 2$. Say a probability measure Γ on E_d is d-invariant if, for any bounded measurable function g on E_{d-1},

$$\int g\big((x_0,\ldots,x_{d-2}),\omega\big)\Gamma(d\mathbf{x},d\omega) = \int g\big((x_1,\ldots,x_{d-1}),T\omega\big)\Gamma(d\mathbf{x},d\omega).$$

Let Φ_d be the distribution of $(X_0,\ldots,X_{d-1},\Theta_0)$ on E_d under $\mathbb{P}_\Phi$. The distribution Φ_d is d-invariant by the shift-invariance of $\mathbb{P}_\Phi$ and by the fact that $\Theta_1 = T\Theta_0$, $\mathbb{P}_\Phi$-almost surely. Now define $I^{(d)}\colon \mathcal{M}_1(E_d) \to [0,\infty]$ by

$$(3.5)\quad I^{(d)}(\Gamma) = \begin{cases} H(\Gamma \mid \Gamma\mathbb{P}^{(d)}), & \text{if } \Gamma \text{ is } d\text{-invariant and has } \Omega\text{-marginal } \pi, \\ \infty, & \text{otherwise.} \end{cases}$$

3.6 Theorem. *The mapping $I^{(d)}$ is lower semicontinuous and convex. Assume* (A). *Then $I^{(d)}(\Gamma) = 0$ if and only if $\Gamma = \Phi_d$, $I^{(d)}$ has compact level sets and $I^{(d)}$ governs the large deviations of $\mathbf{M}_n^{(d)}$ under $\mathbb{P}_{(x,\omega)}$, uniformly over $\mathcal{X}$ and π-almost surely.*

We are ready to pass to process level. Write $\xi = (x_0, x_1, x_2, \ldots)$ for elements of $\mathcal{X}^{\mathbb{Z}^+}$, and $\zeta = (\xi,\omega)$ for elements of E_∞. Define a continuous map U on E_∞ by $U\zeta = (S\xi, T\omega)$, and let $\mathcal{M}_U(E_\infty)$ denote the space of U-invariant probability measures on E_∞. Let the probability measure Φ_∞ on E_∞ be the distribution of $(\vec{X}_0, \Theta_0)$ under $\mathbb{P}_\Phi$.

Let Θ denote the Ω-valued projection on E_∞. Let $\mathcal{E}(n)$ be the σ-field generated by $(X_0, \dots, X_{n-1}, \Theta)$. For $\Gamma \in \mathcal{M}_U(E_\infty)$, define

$$h(\Gamma \mid \Phi_\infty) = \lim_{n\to\infty} \frac{1}{n} H_{\mathcal{E}(n)}(\Gamma \mid \Phi_\infty), \tag{3.7}$$

assuming for the moment that the limit exists, and then define $I^{(\infty)}\colon \mathcal{M}_1(E_\infty) \to [0, \infty]$ by

$$I^{(\infty)}(\Gamma) = \begin{cases} h(\Gamma \mid \Phi_\infty), & \text{if } \Gamma \text{ is } U\text{-invariant}, \\ \infty, & \text{otherwise}. \end{cases} \tag{3.8}$$

3.9 THEOREM. *Assume* (A). *Then Φ_∞ is U-invariant and U-ergodic, the limit in* (3.7) *exists for all $\Gamma \in \mathcal{M}_U(E_\infty)$, and $h(\Gamma \mid \Phi_\infty)$ can be finite only if Γ has Ω-marginal π. The functional $h(\cdot \mid \Phi_\infty)$ is an affine function on $\mathcal{M}_U(E_\infty)$ and $I^{(\infty)}$ is lower semicontinuous, is convex, has compact level sets and $I^\infty(\Gamma) = 0$ if and only if $\Gamma = \Phi_\infty$.*

Moreover, I^∞ governs the large deviations of $\mathbf{M}_n$ under $\mathbb{P}_{(x,\omega)}$, uniformly over $\mathcal{X}$ and π-almost surely.

The proofs of Section 4 show that our definitions for $\mathbf{M}_n^{(d)}$ and $\mathbf{M}_n$ are natural for the setting. The more standard definitions of multivariate empirical distributions and empirical processes are

$$\mathbf{L}_n^{(d)} = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{(Z_k, \dots, Z_{k+d-1})},$$

with values in $\mathcal{M}_1(E^d)$, and

$$\mathbf{R}_n = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{\vec{Z}_k},$$

with values in $\mathcal{M}_1(E^{\mathbb{Z}^+})$, where

$$\vec{Z}_k = \left(\vec{X}_k, \vec{\Theta}_k\right) = \left((X_k, X_{k+1}, X_{k+2}, \dots), (\Theta_k, \Theta_{k+1}, \Theta_{k+2}, \dots)\right).$$

Let us also record the large deviation principles of these random variables.

Define the shift-invariant probability measure $\overline{\pi}$ on $\Omega^{\mathbb{Z}^+}$ by the rule

$$\overline{\pi}\left\{(\Theta_0, \dots, \Theta_n) \in B\right\} = \int \mathbf{1}_B(\omega, T\omega, \dots, T^n\omega)\pi(d\omega).$$

For $d \in \mathbb{N}$, the marginal distribution of $(\Theta_0, \dots, \Theta_{d-1})$ is denoted by $\overline{\pi}_d$. Elements of E^d are written $\mathbf{z} = (z_0, \dots, z_{d-1})$.

3.10 DEFINITION. Say a probability measure μ on E^d is *shift-invariant* if, for any bounded measurable function g on E^{d-1},

$$\int_{E^d} g(z_0,\dots,z_{d-2})\mu(d\mathbf{z}) = \int_{E^d} g(z_1,\dots,z_{d-1})\mu(d\mathbf{z}).$$

It is immediate that a probability measure Q on $E^{\mathbb{Z}^+}$ is shift-invariant if and only if Q_d is shift-invariant for all d, where Q_d is the marginal distribution of $(Z_0,\dots,Z_{d-1})$ under Q.

For $\mu \in \mathcal{M}_1(E^d)$, let μ_{d-1} denote the marginal distribution of $(Z_0,\dots,Z_{d-2})$ under μ. The probability measure $\mu_{d-1} \otimes \mathbb{P}$ on E^d is then defined by

$$\mu_{d-1} \otimes \mathbb{P}(C) = \iint \mathbf{1}_C(z_0,\dots,z_{d-1})\mathbb{P}(z_{d-2},dz_{d-1})\mu_{d-1}(d\mathbf{z}'), \tag{3.11}$$

where $\mathbf{z}' = (z_0,\dots,z_{d-2})$. The d-variate rate is defined for $\mu \in \mathcal{M}_1(E^d)$ by

$$K^{(d)}(\mu) = \begin{cases} H(\mu \mid \mu_{d-1} \otimes \mathbb{P}), & \text{if } \mu \text{ is shift-invariant and} \\ & \quad \text{has marginal } \overline{\pi}_d \text{ on } \Omega^d, \\ \infty, & \text{otherwise.} \end{cases} \tag{3.12}$$

3.13 THEOREM. *Let $d \geq 2$ and assume* (A). *Then $K^{(d)}$ is lower semicontinuous, is convex, has compact level sets and $K^{(d)}(\mu) = 0$ if and only if μ is the E^d-marginal of $\mathbb{P}_\Phi$. Moreover, $K^{(d)}$ governs the large deviations of $\mathbf{L}_n^{(d)}$ under $\mathbb{P}_{(x,\omega)}$, uniformly over $\mathcal{X}$ and π-almost surely.*

Let $\mathcal{F}(n,Z)$ denote the σ-field generated by $(Z_0,\dots,Z_{n-1})$. For a shift-invariant probability measure Q on $E^{\mathbb{Z}^+}$, the specific entropy of Q relative to $\mathbb{P}_\Phi$ is given by

$$h(Q \mid \mathbb{P}_\Phi) = \lim_{n\to\infty} \frac{1}{n} H_{\mathcal{F}(n,Z)}(Q \mid \mathbb{P}_\Phi). \tag{3.14}$$

For $Q \in \mathcal{M}_1(E^{\mathbb{Z}^+})$ define

$$K^{(\infty)}(Q) = \begin{cases} h(Q \mid \mathbb{P}_\Phi), & \text{if } Q \text{ is shift-invariant,} \\ \infty, & \text{otherwise.} \end{cases} \tag{3.15}$$

3.16 THEOREM. *Assume* (A). *Then the limit in* (3.14) *exists for all shift-invariant probabilities Q, and $h(Q \mid \mathbb{P}_\Phi)$ can be finite only if Q has marginal $\overline{\pi}$ on $\Omega^{\mathbb{Z}^+}$. The functional $K^{(\infty)}$ is lower semicontinuous, is convex, is affine on the space of shift-invariant probabilities, has compact level sets and $K^{(\infty)}(Q) = 0$ if and only if $Q = \mathbb{P}_\Phi$. Moreover, $K^{(\infty)}$ governs the large deviations of $\mathbf{R}_n$ under $\mathbb{P}_{(x,\omega)}$, uniformly over $\mathcal{X}$ and π-almost surely.*

Theorems 3.13 and 3.16 can be compared with some earlier results. Consider the following hypothesis: There are an integer b and a constant M such that, for all $z, z' \in E$ and $C \in \mathcal{B}_E$,

$$\mathbb{P}^b(z, C) \leq M\mathbb{P}^b(z', C). \tag{3.17}$$

Under this hypothesis, Theorems 1.2 and 1.4 of Ellis (1988) and Theorem 1.3 of Ellis and Wyner (1989) state large deviation principles for $\mathbf{L}_n$, $\mathbf{L}_n^{(d)}$ and $\mathbf{R}_n$. The respective rates are as in (3.2), (3.12) and (3.15), but without the provisions concerning the Ω-marginals. Thus their process level rate agrees with ours, but the position and multivariate position level rates disagree. This appears inconsistent, for the position rates can be expressed in terms of the process rate, by the push-forward principle. How can they differ if they come from the same process rate? The answer is that (3.17) forces Ω to be a singleton and the conditions on Ω-marginals become vacuous.

The related hypothesis (U) of Deuschel and Stroock [(1989), page 100] requires that there exist b, N and M such that

$$\mathbb{P}^b(z, C) \leq \frac{M}{N} \sum_{k=1}^{N} \mathbb{P}^k(z', C),$$

for all $z, z' \in E$ and $C \in \mathcal{B}_E$. Hypothesis (U) neither implies Assumption (A) nor is implied by (A). Hypothesis (U) forces Ω to be finite, which Assumption (A) does not. On the other hand, taking $\mathcal{X} = \Omega = \{0, 1\}$, taking T to be the flip on Ω and taking two transition kernels

$$P(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad P(1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

on $\mathcal{X}$ gives a simple example of a setting satisfying (U) but not Assumption (A).

3.2. *Sample chains.* For the sample chain, the empirical distributions are

$$\mathbf{L}_n^X = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k}$$

and the empirical processes

$$\mathbf{R}_n^X = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{\vec{X}_k},$$

with values in $\mathcal{M}_1(\mathcal{X})$ and $\mathcal{M}_1(\mathcal{X}^{\mathbb{Z}^+})$, respectively. For probabilities μ on $\mathcal{X}$, define

$$I_X(\mu) = \inf\{\mathbb{J}(\Gamma) : \Gamma \in \mathcal{M}_1(E),\ \Gamma \text{ has marginals } \mu \text{ and } \pi\}. \tag{3.18}$$

Let the probability measure φ on $\mathcal{X}$ be the marginal of Φ.

3.19 THEOREM. *Assume* (A). *Then I_X is lower semicontinuous, is convex, has compact level sets and $I_X(\mu) = 0$ if and only if $\mu = \varphi$. The functional I_X governs* *the large deviations of* $\mathbf{L}_n^X$ *under* P_x^ω, *uniformly over* $\mathcal{X}$ *and* π*-almost surely, in the following sense: There is a Borel subset* Ω_X *of* Ω *such that* $\pi(\Omega_X) = 1$ *and, for all* $\omega \in \Omega_X$, *these large deviation bounds hold: If* $F \subset \mathcal{M}_1(\mathcal{X})$ *is closed, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathcal{X}} P_x^\omega\{\mathbf{L}_n^X \in F\} \le -I_X(F),$$

and if $G \subset \mathcal{M}_1(\mathcal{X})$ *is open, then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in\mathcal{X}} P_x^\omega\{\mathbf{L}_n^X \in G\} \ge -I_X(G).$$

For $f \in C_b(\mathcal{X})$, let

$$S_n f = \sum_{k=0}^{n-1} f(X_k),$$

and define

$$c(f) = \inf_{n\in\mathbb{N}} \frac{1}{n} \int \left[\log \sup_{x\in\mathcal{X}} \int \exp(S_n f)\, dP_x^\omega\right] \pi(d\omega). \tag{3.20}$$

3.21 THEOREM. *As a functional on* $C_b(\mathcal{X})$, c *is Lipschitz continuous and convex. Under Assumption* (A), c *and* I_X *are in duality*:

$$c(f) = \sup\left\{\int f\, d\mu - I_X(\mu) : \mu \in \mathcal{M}_1(\mathcal{X})\right\}$$

and

$$I_X(\mu) = \sup\left\{\int f\, d\mu - c(f) : f \in C_b(\mathcal{X})\right\}.$$

3.22 EXAMPLE (Independent variables). Suppose we are in the canonical setup described in Example 2.4, and let π_0 be the marginal of π on $\mathcal{P}(\mathcal{X})$. Assume that π-almost every transition $P(\omega; x, dy)$ is independent of the starting state x, in other words, that the variables X_n are conditionally independent given $\omega \in \Omega$. Then we may think of π_0 as a probability measure on $\mathcal{M}_1(\mathcal{X})$, and we have

$$c(f) = \int_{\mathcal{M}_1(\mathcal{X})} \left[\log \int_{\mathcal{X}} e^f\, d\nu\right] \pi_0(d\nu).$$

The rate can be written

$$I_X(\mu) = \sup\left\{\iint \log \frac{u(x)}{\int u\, d\nu} \mu(dx) \pi_0(d\nu) : u \in V_b(\mathcal{X})\right\}. \tag{3.23}$$

Note the curious fact that the rate depends on π only through the marginal π_0. That this is not true in general is demonstrated by Example 3.27.

3.24 REMARK. Equation (3.23) bears a pleasant similarity to the Donsker–Varadhan functional, so one would like to know whether the formula generalizes to the nonindependent case. Set

$$K(\mu) = \sup\left\{ \iint \log \frac{u(x)}{P(\omega;\, x, u)} \mu(dx)\pi(d\omega) : u \in V_b(\mathcal{X}) \right\}.$$

Following the proof of Deuschel and Stroock's (1989) Lemma 4.1.45, it is not hard to see that $K(\mu) = 0$ if and only if μ is invariant for $\mathbf{P}$, the mean transition defined below. Thus the zeros of K and I_X do not necessarily coincide, and K cannot in general represent the rate (see Example 3.27).

Our description of the rate I_X is somewhat indirect, so let us investigate it further by finding upper and lower bounds in terms of familiar functions. Define, for $\mu \in \mathcal{M}_1(\mathcal{X})$,

$$J_\pi(\mu) = \int_\Omega J_\omega(\mu)\pi(d\omega),$$

where J_ω is the Donsker–Varadhan functional of the transition kernel $P(\omega)$; J_ω is a lower semicontinuous function of ω, so measurability is not a problem in the above integral.

3.25 THEOREM. *$I_X(\mu) \le J_\pi(\mu) = \mathbb{J}(\mu \otimes \pi)$ for all $\mu \in \mathcal{M}_1(\mathcal{X})$.*

For a natural lower bound, we need to make a further assumption:

ASSUMPTION (B). *The pair (Ω, T) is as in the canonical setting of Example 2.4, and $\pi = \pi_0^{\mathbb{Z}}$ for some Borel probability measure π_0 on $\mathcal{P}(\mathcal{X})$.*

Define the mean transition $\mathbf{P}$ on $\mathcal{X}$ by

$$\mathbf{P}(x, A) = \int_\Omega P(\omega;\, x, A)\pi(d\omega),$$

with the corresponding Donsker–Varadhan functional

$$\mathbf{J}(\mu) = \sup\left\{ \int \log \frac{u}{\mathbf{P}u}\, d\mu : u \in V_b(\mathcal{X}) \right\}.$$

Under Assumption (B), the mean process

$$\mathbf{P}_x = \int_\Omega P_x^\omega \pi(d\omega)$$

is Markovian with transition probabilities $\mathbf{P}(x, dy)$.

3.26 THEOREM. *Assume* (A) *and* (B). *Then* $\mathbf{J}(\mu) \le I_X(\mu)$ *for all* $\mu \in \mathcal{M}_1(\mathcal{X})$.

3.27 EXAMPLE. Let us see how the above theorem can fail when Assumption (B) is not in force and the zeros of I_X and $\mathbf{J}$ do not coincide. Take $\mathcal{X} = \{0, 1\}$, $\Omega = \mathcal{P}(\mathcal{X})^{\mathbb{Z}}$, T equal to the shift map on Ω and $\varepsilon \in (0,1)$, and define

$$P_0 = \begin{pmatrix} \varepsilon & 1-\varepsilon \\ 1-\varepsilon & \varepsilon \end{pmatrix} \quad \text{and} \quad P_1 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

Let us put the following two ergodic measures on Ω: The fair coin-tossing measure

$$\pi = \left((\delta_{P_0} + \delta_{P_1})/2\right)^{\mathbb{Z}},$$

and the Markovian measure

$$\pi' = \left(\delta_{\omega^0} + \delta_{T\omega^0}\right)/2,$$

where $\omega^0 \in \Omega$ is defined by $\omega^0_k = P_{k \bmod 2}$, $k \in \mathbb{Z}$. Clearly, Assumption (A) holds with $b = 1$ and $M \equiv \varepsilon^{-1} \vee (1-\varepsilon)^{-1}$, for both π and π'; π and π' also have identical marginals on $\mathcal{P}(\mathcal{X})$. Let φ and φ' be the $\mathcal{X}$-marginals of the invariant measures of the skew processes corresponding to the background measures π and π', respectively. These are found to be

$$\varphi = \left(\frac{1-\varepsilon}{3-2\varepsilon}, \frac{2-\varepsilon}{3-2\varepsilon}\right) \quad \text{and} \quad \varphi' = \left(\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}\right).$$

If $\varepsilon = \frac{1}{2}$, the variables X_n are conditionally independent given $\omega \in \Omega$. Assume $\varepsilon \ne \frac{1}{2}$. Then φ and φ' are different, so, in particular, the rates I_X and I'_X corresponding to π and π' must be different. The mean transition for both π and π' is

$$\mathbf{P} = \begin{pmatrix} \varepsilon/2 & 1-\varepsilon/2 \\ (1-\varepsilon)/2 & (1+\varepsilon)/2 \end{pmatrix};$$

φ is $\mathbf{P}$-invariant but φ' is not. Thus $\mathbf{J}(\varphi) = 0 < I'_X(\varphi)$ and $\mathbf{J}(\varphi') > 0 = I'_X(\varphi')$.

Let us move to the process level of sample chains. For shift-invariant probabilities Q on $\mathcal{X}^{\mathbb{Z}^+}$, set

$$h_X(Q) = \inf h\left(\Gamma \mid \Phi_\infty\right), \tag{3.28}$$

where the infimum is over U-invariant probabilities Γ on E_∞ with marginals Q and π. The process level rate function on $\mathcal{M}_1(\mathcal{X}^{\mathbb{Z}^+})$ is given by

$$I_X^{(\infty)}(Q) = \begin{cases} h_X(Q), & \text{if } Q \text{ is shift-invariant,} \\ \infty, & \text{otherwise.} \end{cases} \tag{3.29}$$

Let the probability measure φ_∞ on $\mathcal{X}^{\mathbb{Z}^+}$ be the marginal of Φ_∞.

3.30 THEOREM. *Assume* (A). *Then* $I_X^{(\infty)}$ *is lower semicontinuous, is convex, has compact level sets and* $I_X^{(\infty)}(Q) = 0$ *if and only if* $Q = \varphi_\infty$. *Moreover,* $I_X^{(\infty)}$ *governs the large deviations of* $\mathbf{R}_n^X$ *under* P_x^ω, *uniformly over* $\mathcal{X}$ *and* π*-almost surely.*

3.31 EXAMPLE. A natural question is whether Assumption (A) guarantees large deviation principles under the ergodic probability φ_∞ on $\mathcal{X}^{\mathbb{Z}}$. The answer is no: Take $(\Omega, T) = (\{-1,+1\}^{\mathbb{Z}}$, shift map) and $\mathcal{X} = \{-1,+1\}$. Let μ be the probability measure on $\mathcal{X}^{\mathbb{Z}} = \Omega$ constructed in Orey and Pelikan [(1988), Example 4.2] that fails to satisfy a large deviation principle. Set $\pi = \mu$, and define the transition kernel $P(\omega)$ on $\mathcal{X}$ by $P(\omega; x, A) = \mathbf{1}_A(\omega_0)$, for $\omega = (\omega_k) \in \Omega$. Assumption (A) is trivially satisfied, but $\varphi_\infty = \mu$ and hence cannot satisfy a large deviation principle.

As in Theorem 3.21, we could write the process rate as a convex dual, but (3.20) does not necessarily define the correct functional on functions that depend on more than one coordinate. We shall develop an expression that mimics the limit of a specific entropy. For the remainder of this section, Q is a fixed shift-invariant probability measure on $\mathcal{X}^{\mathbb{Z}^+}$. For $n \in \mathbb{N}, \mathcal{F}(n,X)$ denotes the σ-field generated by $(X_0,\dots,X_{n-1})$, and $\mathcal{C}_n$ denotes the space of bounded, continuous, $\mathcal{F}(n,X)$-measurable functions on $\mathcal{X}^{\mathbb{Z}^+}$. Define

$$K_n(Q) = \sup\left\{ \int f\,dQ - \int \left[\log \int e^f\,dP^\omega_{\varphi^\omega}\right]\pi(d\omega) : f \in \mathcal{C}_n \right\}, \tag{3.32}$$

where $P^\omega_{\varphi^\omega}$ is the sample chain with initial distribution φ^ω and the initial state of the dynamical system at ω, and $\varphi^\omega(dx)$ is a conditional distribution of Φ.

3.33 THEOREM. *Under Assumption* (A), $h_X(Q) = \lim_{n\to\infty} K_n(Q)/n$.

3.34 REMARK. If we assume (A) and choose to ignore the b first coordinates, the above convergence is uniform over initial distributions. For $\mu \in \mathcal{M}_1(\mathcal{X})$, define

$$K_{\mu,n}(Q) = \sup\left\{ \int f\,dQ - \int \left[\log \int \exp\left(f \circ S^b\right) dP^\omega_\mu\right]\pi(d\omega) : f \in \mathcal{C}_n \right\}.$$

Then minor modifications in the proof of Theorem 3.33 (Section 5) show that

$$\limsup_{n\to\infty} \sup_{\mu\in\mathcal{M}_1(\mathcal{X})} \frac{1}{n} K_{\mu,n}(Q) \le h_X(Q) \le \liminf_{n\to\infty} \inf_{\mu\in\mathcal{M}_1(\mathcal{X})} \frac{1}{n} K_{\mu,n}(Q).$$

Finally, let us develop bounds for $h_x(Q)$ in terms of functions that resemble familiar specific entropies. Define

$$\overline{h}(Q \mid \varphi_\infty) = \limsup_{n\to\infty} \frac{1}{n} H_{\mathcal{F}(n,X)}(Q \mid \varphi_\infty) \tag{3.35}$$

and

$$h_\pi(Q) = \lim_{d\to\infty} \int H\big(Q_d \mid Q_{d-1} \otimes P(\omega)\big)\pi(d\omega). \tag{3.36}$$

The limit in (3.35) is not guaranteed to exist, hence the lim sup; Q_d is the marginal of Q on $\mathcal{F}(d,X)$, and the meaning of $Q_{d-1}\otimes P(\omega)$ is analogous to (3.11). The limit in (3.36) exists by the monotonicity of relative entropy.

3.37 THEOREM. *Under Assumption* (A), $\overline{h}(Q \mid \varphi_\infty) \le h_X(Q) \le h_\pi(Q) = h(Q \otimes \pi \mid \Phi_\infty)$. *Under Assumptions* (A) *and* (B), *we have* $\varphi_\infty = \mathbf{P}_\varphi$, *so the lower bound reads* $\overline{h}(Q \mid \mathbf{P}_\varphi) \le h_X(Q)$.

4. Proofs of the basic properties. Let us begin with a simple observation about compact-open topologies.

4.1 LEMMA. *Let E and F be Hausdorff spaces with countable bases for their topologies, and suppose that E is locally compact. Then the compact-open topology of $C(E,F)$ has a countable base.*

PROOF. For $H \subset E$ and $V \subset F$, let $S(H,V) = \{f \in C(E,F): f(H) \subset V\}$. By definition, the compact-open topology is generated by the subbase

$$\mathcal{S} = \{S(H,V): H \subset E \text{ is compact, and } V \subset F \text{ is open}\};$$

see Munkres [(1975), page 286]. Let $\mathcal{K}$ and $\mathcal{U}$ be countable bases of open sets for the topologies of E and F, respectively, and furthermore so that the closure $\overline{K}$ is compact for each $K \in \mathcal{K}$. Let

$$\mathcal{S}_0 = \left\{S\left(\overline{K}, \bigcup_{j=1}^{r} U_j\right): K \in \mathcal{K},\ r \in \mathbb{N}, U_1, \dots, U_r \in \mathcal{U}\right\}.$$

The class $\mathcal{S}_0$ is a countable subcollection of $\mathcal{S}$. By Munkres [(1975), Lemma 8.2 in Chapter 3] the collection of finite intersections of elements of $\mathcal{S}_0$ gives a countable base for the compact-open topology of $C(E,F)$. □

PROOF OF LEMMA 2.2. Easy arguments show that D is compatible with the compact-open topology, and then separability follows from Lemma 4.1. The completeness of D follows from the completeness of $(\mathcal{M}_1(\mathcal{X}), r)$ and the fact that a uniform limit of continuous functions is itself continuous. □

PROOF OF LEMMA 2.3. Let $P(\omega; x)$ be the value of the map $P(\omega): \mathcal{X} \to \mathcal{M}_1(\mathcal{X})$ at x. The composition $(x,\omega) \mapsto (x, P(\omega)) \mapsto P(\omega; x)$ from E into $\mathcal{M}_1(\mathcal{X})$ is continuous, the second step by Theorem 5.3 of Munkres [(1975), page 287]. Since $\mathbb{P}: E \to \mathcal{M}_1(E)$ can be expressed by $\mathbb{P}(x,\omega) = P(\omega; x) \otimes \delta_{T\omega}$, it is clearly continuous. □

4.2 LEMMA. *Assume* (A). *Let $A \in \mathcal{B}_E$ and $x \in \mathcal{X}$. Then there is a Borel subset Ω_1 of Ω, depending on x and A, such that $\pi(\Omega_1) = 1$ and the following inequality holds for all $\omega \in \Omega_1$ and $y \in \mathcal{X}$:*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{P}^n\big((y,\omega), A\big) \ge \int_\Omega M(\eta)^{-1} \mathbb{P}^b\big((x,\eta), A\big)\pi(d\eta).$$

PROOF. For $n \geq b, y \in \mathcal{X}$ and $\omega \in \Omega_A$,

$$\mathbb{P}^n\big((y,\omega), A\big) \geq M(T^{n-b}\omega)^{-1}\mathbb{P}^b\big((x, T^{n-b}\omega), A\big),$$

so, for $N \geq b$,

$$\frac{1}{N}\sum_{n=1}^{N}\mathbb{P}^n\big((y,\omega), A\big) \geq \frac{1}{N}\sum_{m=0}^{N-b} M(T^m\omega)^{-1}\mathbb{P}^b\big((x, T^m\omega), A\big).$$

Now let $N \to \infty$ and use the pointwise ergodic theorem. □

To understand the consequences of Assumption (A) for the ergodic behavior of the skew chain, we shall apply the theory of Hopf Markov chains, as presented by Foguel (1969). A *Hopf Markov chain* is a quadruple $(Z, \mathcal{Z}, \nu, P)$, where $(Z, \mathcal{Z}, \nu)$ is a σ-finite measure space and P is a positive contraction on $L^1(\nu)$. The action of P on an $L^1(\nu)$-function u is written uP. If, instead of an $L^1(\nu)$-contraction, we are given a Markov transition kernel P on the measure space $(Z, \mathcal{Z}, \nu)$ satisfying

(4.3) $\quad P(z, A) = 0$ for ν-almost all z, whenever $A \in \mathcal{Z}$ and $\nu(A) = 0$,

we can define a positive $L^1(\nu)$-contraction as follows: Given $u \in L^1(\nu)$, let β be the finite signed measure defined by $d\beta = u\,d\nu$. Define a new finite signed measure βP by

$$\beta P(A) = \int P(z, A)\beta(dz),$$

for $A \in \mathcal{Z}$. By (4.3), $\beta P \ll \nu$, and we set $uP = d(\beta P)/d\nu$.

We shall need the following standard fact, whose proof can be found in Rosenblatt (1971):

4.4 LEMMA. *Suppose $P(x, A)$ is a Markov transition kernel on a measurable space $(X, \mathcal{A})$ and that Φ is an invariant probability for P on $(X, \mathcal{A})$. Let P_Φ be the shift-invariant probability measure on $(X^{\mathbb{Z}^+}, \mathcal{A}^{\mathbb{Z}^+})$ corresponding to the Markov chain with initial distribution Φ and transition probabilities $P(x, dy)$. Let S denote the shift on $X^{\mathbb{Z}^+}$, and define the following sub-σ-fields of $\mathcal{A}$ and $\mathcal{A}^{\mathbb{Z}^+}$:*

$$\mathcal{E} = \{A \in \mathcal{A}: P\mathbf{1}_A = \mathbf{1}_A,\ \Phi\text{-}a.s.\};$$

$$\mathcal{I} = \{B \in \mathcal{A}^{\mathbb{Z}^+}: B = S^{-1}B,\ P_\Phi\text{-}a.s.\}.$$

Then $\Phi(A) = 0$ or 1 for all $A \in \mathcal{E}$ if and only if $P_\Phi(B) = 0$ or 1 for all $B \in \mathcal{I}$, that is, if and only if the coordinate process is ergodic under P_Φ.

Now fix $\widehat{x} \in \mathcal{X}$ and define a probability Π for Borel subsets C of E by

$$\Pi(C) = \int \mathbb{P}^b\big((\widehat{x}, \omega), C\big)\pi(d\omega).$$

4.5 LEMMA. *Assume* (A). *Then* $(E, \mathcal{B}_E, \Pi, \mathbb{P})$ *is a Hopf Markov chain. There is a unique probability* Φ *on* E *that is both* $\mathbb{P}$*-invariant and absolutely continuous with respect to* Π. *The* Ω*-marginal* $\Phi_\Omega = \pi$, *and* $\mathbb{P}_\Phi$ *is ergodic.*

PROOF. First we check the analogue of (4.3). Suppose $C \in \mathcal{B}_E$ is such that $\Pi(C) = 0$. Then $\mathbb{P}^b((\widehat{x}, T\omega), C) = 0$ for π-almost all ω, and

$$\int \mathbb{P}(z, C)\Pi(dz) \le \int M(T\omega)\mathbb{P}^b\big((\widehat{x}, T\omega), C\big)\pi(d\omega) = 0.$$

To prove the existence of an invariant probability, absolutely continuous with respect to Π, it suffices to show that we cannot have an increasing sequence $\{A_j\}$ of Borel sets such that both $A_j \nearrow E$, Π-almost surely, and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{P}^n(z, A_j) = 0 \quad \text{for } \Pi\text{-almost all } z \text{ and all } j,$$

according to Foguel [(1969), Corollary 2 on page 46].

Suppose we have such sets $\{A_j\}$ and let A be their union. By Lemma 4.2,

$$0 = \lim_{j\to\infty}\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{P}^n\big((y,\omega), A_j\big) \ge \int_\Omega M(\eta)^{-1}\mathbb{P}^b\big((\widehat{x}, \eta), A\big)\pi(d\eta),$$

for Π-almost all $(y, \omega) \in E$, so that $\Pi(A) = 0$, a contradiction. This proves the existence of a probability Φ on E such that $\Phi \ll \Pi$ and $\Phi\mathbb{P} = \Phi$.

Since π and Φ_Ω are T-invariant, so is the density $d\Phi_\Omega/d\pi$. By the ergodicity of π, $d\Phi_\Omega/d\pi = 1$ π-almost surely, hence $\Phi_\Omega = \pi$.

To prove ergodicity of $\mathbb{P}_\Phi$, let $A \in \mathcal{B}_E$ be such that $\mathbb{P}\mathbf{1}_A = \mathbf{1}_A$, Φ-almost surely. This and Lemma 4.2 give

$$\mathbf{1}_A = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{P}^n\mathbf{1}_A \ge \int_\Omega M(\eta)^{-1}\mathbb{P}^b\big((\widehat{x}, \eta), A\big)\pi(d\eta),$$

Φ-almost surely, since $\Phi_\Omega = \pi$. Multiply by $\mathbf{1}_{A^c}$ and integrate to get

$$0 \ge \Phi(A^c)\int_\Omega M(\eta)^{-1}\mathbb{P}^b\big((\widehat{x}, \eta), A\big)\pi(d\eta).$$

If $\Phi(A^c) > 0$, then $\mathbb{P}^b\big((\widehat{x}, \eta), A\big) = 0$ for π-almost all η, from which it follows that $\Pi(A) = 0$ and, consequently, $\Phi(A) = 0$. We have shown that $\Phi(A)$ is 0 or 1, and so, by Lemma 4.4, $\mathbb{P}_\Phi$ is ergodic.

Suppose that Γ is another $\mathbb{P}$-invariant probability such that $\Gamma \ll \Pi$. Then the same holds for $\Psi = (\Phi + \Gamma)/2$, too. By the above reasoning, $\mathbb{P}_\Gamma$ and $\mathbb{P}_\Psi$ are ergodic, but then $\mathbb{P}_\Psi = (\mathbb{P}_\Phi + \mathbb{P}_\Gamma)/2$ forces $\Psi = \Phi = \Gamma$. □

PROOF OF THEOREM 2.8. Let Φ be the probability measure whose existence was derived in the above lemma. For part (i) of Theorem 2.8, it remains to prove uniqueness. Let Γ be $\mathbb{P}$-invariant and satisfy $\Gamma_\Omega = \pi$. By the above lemma, it suffices to show that $\Gamma \ll \Pi$. So let $C \in \mathcal{B}_E$ be such that $\Pi(C) = 0$. Then

$$\Gamma(C) = \Gamma\mathbb{P}^b(C) = \int \mathbb{P}^b\big((x,\omega), C\big)\Gamma(dx, d\omega)$$
$$\leq \int M(\omega)\mathbb{P}^b\big((\widehat{x},\omega), C\big)\pi(d\omega) = 0.$$

Part (ii) was already proved in Lemma 4.5.

For part (iii), let D be the Borel subset of $E^{\mathbb{Z}^+}$ where the convergence

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(Z_k) = \int f\,d\Phi$$

takes place. Since D is a tail event, (2.7) implies that for each $\omega \in \Omega_A$ either $\mathbb{P}_{(x,\omega)}(D^c) = 0$, for all $x \in \mathcal{X}$, or $\mathbb{P}_{(x,\omega)}(D) = 0$, for all $x \in \mathcal{X}$. By ergodicity $\mathbb{P}_{(x,\omega)}(D^c) = 0$ for Φ-almost all (x,ω), so it must be that $\mathbb{P}_{(x,\omega)}(D^c) = 0$ for all x and π-almost all ω. □

5. Proofs of the large deviation theorems. We shall first prove the large deviation theorems for the skew chain, for the sample chain results will then follow by the push-forward principle. For the skew chain, we proceed from position level to process level via multivariate position level results.

To begin, we present some standard facts about large deviations and relative entropy that will be used repeatedly. Proofs can be found in Deuschel and Stroock (1989) and Varadhan (1984). The *push-forward* or *contraction principle* says that if I governs the large deviations of $\{\xi_n\}$ and has compact level sets, and if f is a continuous map from $\mathcal{S}$ into another metric space $\mathcal{T}$, then the function J, defined on $\mathcal{T}$ by $J(y) = \inf\{I(x): f(x) = y\}$, also has compact level sets and governs the large deviations of $\{f(\xi_n)\}$.

If μ and ν are Borel measures on a Polish space $\mathcal{X}$, relative entropy can be expressed as

$$H(\mu \mid \nu) = \sup\left\{\int f\,d\mu - \log\int e^f\,d\nu : f \in C_b(\mathcal{X})\right\}. \tag{5.1}$$

An easy consequence is

$$H\big(\varepsilon\mu + (1-\varepsilon)\mu' \mid \varepsilon\nu + (1-\varepsilon)\nu'\big) \leq \varepsilon H(\mu \mid \nu) + (1-\varepsilon)H(\mu' \mid \nu'), \tag{5.2}$$

for probabilities μ, μ', ν and ν' and for $0 \leq \varepsilon \leq 1$. Moreover, $H(\mu \mid \nu)$ is lower semicontinuous as a function of μ. The level sets $\{\mu \in \mathcal{M}_1(\mathcal{X}): H(\mu \mid \nu) \leq l\}$ are compact in the weak topology of $\mathcal{M}_1(\mathcal{X})$, for all $l \in \mathbb{R}$. If $\mathcal{A}$ is a countably generated sub-σ-field of $\mathcal{B}_\mathcal{X}$, and if $\mu^\mathcal{A}$ and $\nu^\mathcal{A}$ are versions of the conditional probabilities of μ and ν, given $\mathcal{A}$, then

$$H(\mu \mid \nu) = H_\mathcal{A}(\mu \mid \nu) + \int H\big(\mu^\mathcal{A} \mid \nu^\mathcal{A}\big)\,d\mu. \tag{5.3}$$

5.1. *Position level for the skew chain.* We shall establish Theorem 3.3 for the shifted empirical distribution

$$\widetilde{\mathbf{L}}_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{Z_{b+k}}.$$

The conclusion for $\mathbf{L}_n$ follows because $\mathbf{L}_n$ and $\widetilde{\mathbf{L}}_n$ come uniformly close as n increases. For a detailed argument, see Orey [(1986), Proposition 3.1].

Our first goal is the upper bound, namely, the following proposition.

5.4 PROPOSITION. *Under Assumption* (A), *there is a Borel subset* Ω_u *such that* $\pi(\Omega_u) = 1$ *and*

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in F\} \le -I_E(F),$$

for all closed $F \subset \mathcal{M}_1(E)$ *and all* $\omega \in \Omega_u$.

For $\omega \in \Omega$, set

$$\mathbf{L}_n(\omega) = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k\omega}$$

and

$$\widetilde{M}(\omega) = \sup_{n\ge 1} \frac{1}{n}\sum_{j=0}^{n-1} \log M(T^j\omega).$$

We let Ω_u be the set of $\omega \in \Omega_A$ such that $\lim_{n\to\infty} \mathbf{L}_n(\omega) = \pi$ in the weak topology of $\mathcal{M}_1(\Omega)$, and $\widetilde{M}(\omega)$ is finite. The set Ω_u is a shift-invariant Borel subset of Ω. By the ergodic theorem, the second countability of $\mathcal{M}_1(\Omega)$ and Assumption (A), Ω_u has π-measure 1.

5.5 LEMMA. *Suppose* $\omega \in \Omega_u$ *and* K *is a compact subset of* $\mathcal{M}_1(E)$, *disjoint from* $\mathcal{M}_\pi(E)$. *Then, for all sufficiently large* n,

$$\sup_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in K\} = 0.$$

PROOF. The image of K under the projection $E \to \Omega$ is a compact subset of $\mathcal{M}_1(\Omega)$, not containing π. Thus $\mathbf{L}_n(T^b\omega)$ lies outside this set for n large enough, and the conclusion follows from the fact that $\mathbb{P}_{(x,\omega)}$-almost surely

$$\frac{1}{n}\sum_{k=b}^{b+n-1} \delta_{\Theta_k} = \mathbf{L}_n(T^b\omega). \qquad \square$$

5.6 LEMMA. *Assume* (A) *and suppose that* $\omega \in \Omega_u$. *Then there are compact sets* $C_l \subset \mathcal{M}_1(E)$, *possibly depending on* ω, *such that, for all* $n, l \in \mathbb{N}$,

$$\sup_{x \in \mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in C_l^c\} \le e^{-nl}.$$

PROOF. Fix $\widehat{x} \in \mathcal{X}$, and define measures $\beta_{n,\omega}$ and β on E by

$$\beta_{n,\omega} = \frac{1}{n}\sum_{j=0}^{n-1} \mathbb{P}^b(\widehat{x}, T^j\omega) \quad \text{and} \quad \beta = \int_\Omega \mathbb{P}^b(\widehat{x}, \eta)\pi(d\eta).$$

Let $\varepsilon_k > 0$ be any numbers such that $\varepsilon_k \searrow 0$ as $k \nearrow \infty$, and define, for $k \in \mathbb{N}$,

$$c_k = \varepsilon_k^{-1}\big(k + 1 + 2\widetilde{M}(\omega) + \log 4\big).$$

Since $T{:}\,\Omega \to \Omega$ and $P{:}\,\Omega \to \mathcal{P}(\mathcal{X})$ are continuous by assumption, $\mathbb{P}^b(\widehat{x}, \omega)$ depends continuously on ω. Therefore the convergence $\mathbf{L}_n(\omega) \to \pi$ guarantees that $\beta_{n,\omega} \to \beta$ as $n \to \infty$, hence $\{\beta_{n,\omega}\}_{n=1}^\infty$ is tight by Prohorov's theorem. Pick compact sets $H_k \subset E$ such that $\beta_{n,\omega}(H_k^c) < \exp(-c_k b)$, for all $n, k \in \mathbb{N}$.

Given n, let $q = q_n$ be an integer such that $(q-1)b < n \le qb$. In the next computation, first add some extra terms to the sum, then apply Chebyshev's and Hölder's inequalities, then Assumption (A) q times, then Jensen's inequality and, finally, the definitions of the various quantities:

$$\begin{aligned}
&\mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n(H_k^c) > \varepsilon_k\} \\
&\quad\le \mathbb{P}_{(x,\omega)}\left\{\frac{1}{n}\sum_{i=1}^{q}\sum_{j=0}^{b-1} \mathbf{1}_{H_k^c}(Z_{ib+j}) > \varepsilon_k\right\} \\
&\quad\le \exp[-nc_k\varepsilon_k] \cdot \prod_{j=0}^{b-1}\left(\int \prod_{i=1}^{q} \exp\left[c_k b \mathbf{1}_{H_k^c}(Z_{ib+j})\right] d\mathbb{P}_{(x,\omega)}\right)^{1/b} \\
&\quad\le \exp[-nc_k\varepsilon_k] \cdot \prod_{j=0}^{b-1}\left(\prod_{i=0}^{q-1} M(T^{ib+j}\omega)\mathbb{P}^b\Big((\widehat{x}, T^{ib+j}\omega), \exp\left[c_k b \mathbf{1}_{H_k^c}\right]\Big)\right)^{1/b} \\
&\quad\le \exp\left[-nc_k\varepsilon_k + \frac{1}{b}\sum_{j=0}^{qb-1} \log M(T^j\omega)\right] \\
&\qquad\quad \times \left(\frac{1}{qb}\sum_{j=0}^{qb-1} \mathbb{P}^b\Big((\widehat{x}, T^j\omega), \exp\left[c_k b \mathbf{1}_{H_k^c}\right]\Big)\right)^q \\
&\quad\le \exp\left[-nc_k\varepsilon_k + q\widetilde{M}(\omega)\right]\Big(\exp(c_k b)\beta_{qb,\omega}(H_k^c) + 1\Big)^q \\
&\quad\le \exp\left[-n(k+1)\right],
\end{aligned}$$

valid for all x, n and k.

For $l \in \mathbb{N}$, put $C_l = \{\nu \in \mathcal{M}_1(E): \nu(H_k^c) \le \varepsilon_k, \text{ for } k \ge l\}$. Each C_l is a compact subset of $\mathcal{M}_1(E)$, and

$$\sup_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in C_l^c\} \le \sum_{k=l}^{\infty} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n(H_k^c) > \varepsilon_k\} \le e^{-nl},$$

for all n and l. □

PROOF OF PROPOSITION 5.4. By Lemma 5.6 and Deuschel and Stroock [(1989), Lemma 2.1.5], we need to prove the upper bound only for compact F. We may also assume that $F \cap \mathcal{M}_\pi(E) \ne \emptyset$, for otherwise the conclusion is immediate from Lemma 5.5. By the definition of I_E, $I_E(F) = \mathbb{J}(F \cap \mathcal{M}_\pi(E))$. Let $c < \mathbb{J}(F \cap \mathcal{M}_\pi(E))$. For each $\alpha \in F \cap \mathcal{M}_\pi(E)$, pick a $u_\alpha \in V_b(E)$ such that

$$\int \log \frac{u_\alpha}{\mathbb{P}u_\alpha}\, d\alpha > c.$$

Since $\mathbb{P}$ is Feller continuous by Lemma 2.3, $\log[u_\alpha(\mathbb{P}u_\alpha)^{-1}]$ is a bounded continuous function on E, and we may find an open neighborhood B_α of α so that

$$\inf_{\mu\in B_\alpha} \int \log \frac{u_\alpha}{\mathbb{P}u_\alpha}\, d\mu > c.$$

Tracing the argument of Donsker and Varadhan [(1975a), pages 8–9] then shows that

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in B_\alpha\} \le -c.$$

Cover the compact set $F \cap \mathcal{M}_\pi(E)$ with a finite union $U = B_{\alpha_1} \cup \cdots \cup B_{\alpha_m}$. The set $F \setminus U$ is a compact subset of $\mathcal{M}_1(E) \setminus \mathcal{M}_\pi(E)$, so, by Lemma 5.5,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in F\} &\le \max_{1\le j\le m} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in B_{\alpha_j}\} \\ &\le -c.\end{aligned}$$

□

Our next goal is the corresponding lower bound.

5.7 PROPOSITION. *Under Assumption* (A), *there is a Borel subset* Ω_l *such that* $\pi(\Omega_l) = 1$ *and*

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{x\in\mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in G\} \ge -I_E(G),$$

for all open $G \subset \mathcal{M}_1(E)$ *and all* $\omega \in \Omega_l$.

We saw above that the definition (3.1) of $\mathbb{J}$ is useful for proving the upper bound, but for the lower bound we need another expression for $\mathbb{J}$. Given a probability measure Γ and a Markov transition kernel $\mathbb{Q}$ on E, we define a probability measure $\Gamma \otimes \mathbb{Q}$ on E^2 by

$$\Gamma \otimes \mathbb{Q}(C) = \iint_{E^2} \mathbf{1}_C(z_0, z_1)\mathbb{Q}(z_0, dz_1)\Gamma(dz_0).$$

By Donsker and Varadhan [(1976), Theorem 2.1],

$$\mathbb{J}(\Gamma) = \inf\{H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}) : \mathbb{Q} \text{ is a Markov kernel with invariant measure } \Gamma\}. \tag{5.8}$$

Recall that Φ is the unique $\mathbb{P}$-invariant probability measure on E with Ω-marginal $\Phi_\Omega = \pi$.

5.9 LEMMA. *Let $\nu \in \mathcal{M}_1(\Omega)$ and $\Gamma \in \mathcal{M}_1(E)$.*

(i) *If $\mathbb{J}(\Gamma) < \infty$, then Γ_Ω is T-invariant.*
(ii) *If $H(\nu \mid \pi) < \infty$ and ν is T-invariant, then $\nu = \pi$.*
(iii) *Assume* (A). *Then there are constants $c_1 > 0$ and $c_2 \in \mathbb{R}$ such that*

$$\mathbb{J}(\Gamma) \geq c_1 H(\Gamma \mid \Phi) + c_2,$$

for all $\Gamma \in \mathcal{M}_\pi(E)$.

PROOF. Statement (i) is immediate; (ii) follows from the ergodicity of π and the fact that two T-invariant probability measures are identical if they agree on T-invariant sets.

Note that the assumption $\Gamma_\Omega = \pi$ is necessary for (iii) to hold. For if Γ is a $\mathbb{P}$-invariant probability measure with $\Gamma_\Omega \neq \pi$, then $\mathbb{J}(\Gamma) = 0$ but $H(\Gamma \mid \Phi) = \infty$ by (ii). The point here is that $\mathbb{J}$ treats all $\mathbb{P}$-invariant probabilities equally, with no preference for those with marginal π.

Write $\varphi^\omega(dx)$ for the conditional distribution of the $\mathcal{X}$-coordinate under Φ, given $\omega \in \Omega$. Let $u \in V_b(E)$. By Assumption (A), for all x and π-almost all ω,

$$\mathbb{P}^b u(x, \omega) \leq M(\omega) \int \mathbb{P}^b u(y, \omega)\varphi^\omega(dy).$$

Now take logarithms, integrate against Γ, apply Jensen's inequality and, finally, use the $\mathbb{P}$-invariance of Φ to get

$$\begin{aligned}\int (\log \mathbb{P}^b u)\, d\Gamma &\leq \int \log M\, d\pi + \int \left[\log \int \mathbb{P}^b u(y, \omega)\varphi^\omega(dy)\right]\pi(d\omega)\\ &\leq \int \log M\, d\pi + \log \int u\, d\Phi.\end{aligned}$$

By Lemma 2.3, $\mathbb{P}^k u \in V_b(E)$ for all k, so, by (3.1),

$$\mathbb{J}(\Gamma) \geq \frac{1}{b}\sum_{k=1}^{b}\int \log \frac{\mathbb{P}^{k-1}u}{\mathbb{P}^k u}\,d\Gamma = \frac{1}{b}\int \log \frac{u}{\mathbb{P}^b u}d\Gamma$$

$$\geq \frac{1}{b}\left\{\int \log u\,d\Gamma - \log\int u\,d\Phi\right\} - \frac{1}{b}\int \log M\,d\pi.$$

Taking the supremum of the last expression over $u \in V_b(E)$ and using (5.1) give

$$\mathbb{J}(\Gamma) \geq \frac{1}{b}H(\Gamma \mid \Phi) - \frac{1}{b}\int \log M\,d\pi. \qquad \square$$

Let $\mathcal{Q}$ denote the set of pairs $(\Gamma, \mathbb{Q})$ such that Γ is a probability measure on E and $\mathbb{Q}$ is a Markov kernel with invariant measure Γ. Write $\mathbb{Q}_\Gamma$ for the shift-invariant measure on $E^{\mathbb{Z}^+}$ obtained by extending the initial distribution Γ with the transition kernel $\mathbb{Q}$. At times we shall consider $\mathbb{Q}_\Gamma$ as a measure on $E^{\mathbb{Z}}$ without inventing new notation for it. $Z_n = (X_n, \Theta_n)$ denotes the coordinate variables on $E^{\mathbb{Z}}$ as well as on $E^{\mathbb{Z}^+}$.

5.10 LEMMA. *Under Assumption* (A), *there exists a countable subcollection* $\mathcal{Q}_0$ *of* $\mathcal{Q}$ *such that, for all open subsets* G *of* $\mathcal{M}_1(E)$,

$$I_E(G) = \inf\{H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}) : \Gamma \in G \text{ and } (\Gamma, \mathbb{Q}) \in \mathcal{Q}_0\},$$

and all pairs $(\Gamma, \mathbb{Q}) \in \mathcal{Q}_0$ *satisfy properties* (i)–(v):

(i) $\Gamma_\Omega = \pi$.
(ii) $H(\Gamma \mid \Phi) < \infty$.
(iii) $H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}) < \infty$.
(iv) $\Gamma \sim \Phi$.
(v) $\mathbb{Q}_\Gamma$ *is ergodic.*

PROOF. Observe first that countability is no restriction by the second countability of the weak topology of $\mathcal{M}_1(E)$. Let $G \subset \mathcal{M}_1(E)$ be open. Combining (3.2) and (5.8), we get

$$I_E(G) = \inf\{H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}) : \Gamma \in G,\ \Gamma_\Omega = \pi \text{ and } (\Gamma, \mathbb{Q}) \in \mathcal{Q}\}, \tag{5.11}$$

where $\inf\emptyset = \infty$ by convention. Since we only need to consider Γ such that $\Gamma_\Omega = \pi$ and $\mathbb{J}(\Gamma) < \infty$, we may restrict attention, by Lemma 5.9(iii), to Γ such that $H(\Gamma \mid \Phi) < \infty$. From this follows $\Gamma \ll \Phi$.

Let $(\Gamma, \mathbb{Q})$ be a candidate in (5.11) satisfying $\Gamma \ll \Phi$. Let $\varepsilon \in (0, 1)$, and define probabilities Q_ε on E^2 and Γ_ε on E by

$$Q_\varepsilon = (1-\varepsilon)\Gamma \otimes \mathbb{Q} + \varepsilon\Phi \otimes \mathbb{P}$$

and

$$\Gamma_\varepsilon = (1-\varepsilon)\Gamma + \varepsilon\Phi.$$

Let $\mathbb{Q}_\varepsilon$ be a transition kernel such that $Q_\varepsilon = \Gamma_\varepsilon \otimes \mathbb{Q}_\varepsilon$, and check that Γ_ε is an invariant measure for $\mathbb{Q}_\varepsilon$. For small enough ε, Γ_ε lies in G and consequently $(\Gamma_\varepsilon, \mathbb{Q}_\varepsilon)$ is also a candidate in (5.11). By (5.2),

$$H(\Gamma_\varepsilon \otimes \mathbb{Q}_\varepsilon \mid \Gamma_\varepsilon \otimes \mathbb{P}) \leq (1-\varepsilon) H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}),$$

so we can do even better in (5.11) by picking $(\Gamma_\varepsilon, \mathbb{Q}_\varepsilon)$ instead of $(\Gamma, \mathbb{Q})$. We have $\Gamma_\varepsilon \sim \Phi$. Thus all but (v) are satisfied.

It remains to show that the stationary Markov chain with initial distribution Γ_ε and transition kernel $\mathbb{Q}_\varepsilon$ is ergodic. Let $A \in \mathcal{B}_E$ be such that $\mathbb{Q}_\varepsilon(z, A) = \mathbf{1}_A(z)$ for Γ_ε-almost all $z \in E$. Since $Q_\varepsilon(A \times A^c) = Q_\varepsilon(A^c \times A) = 0$, it follows that $\Phi \otimes \mathbb{P}(A \times A^c) = \Phi \otimes \mathbb{P}(A^c \times A) = 0$, and from this that $\mathbb{P}\mathbf{1}_A = \mathbf{1}_A$, Φ-almost surely. By the ergodicity of $\mathbb{P}_\Phi$ and Lemma 4.4, $\Phi(A) \in \{0,1\}$, but then $\Gamma_\varepsilon(A) \in \{0,1\}$, too, and we are done. □

5.12 LEMMA. *Suppose* $(\Gamma, \mathbb{Q}) \in \mathcal{Q}_0$. *For* $j, k \in \mathbb{Z}$ *and* $f \in L^1(\mathbb{Q}_\Gamma)$, *the following holds* $\mathbb{Q}_\Gamma$*-almost surely*: $\Theta_k = T^{k-j}\Theta_j$, *the σ-fields generated by* Θ_j *and* Θ_k *are equal, as are the conditional expectations* $\mathbb{Q}_\Gamma(f \mid \Theta_j)$ *and* $\mathbb{Q}_\Gamma(f \mid \Theta_k)$.

PROOF. From

$$\infty > H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}) = \int H\big(\mathbb{Q}(z, \cdot) \mid \mathbb{P}(z, \cdot)\big)\Gamma(dz),$$

it follows that $\mathbb{Q}(z, \cdot) \ll \mathbb{P}(z, \cdot)$ for Γ-almost all z, and hence $\mathbb{Q}\big((x, \omega), \mathcal{X} \times \{T\omega\}\big) = 1$ for Γ-almost all (x, ω). This implies that, for $k \geq 0$, $\Theta_k = T^k\Theta_0$, $\mathbb{Q}_\Gamma$-almost surely. The same conclusion for $k < 0$ comes by the shift-invariance of $\mathbb{Q}_\Gamma$ and the invertibility of T. The lemma follows immediately. □

For $\omega \in \Omega$, define a probability measure $\mathbb{P}_\Phi^\omega$ on $E^{\mathbb{Z}^+}$ by

$$\begin{aligned} &\mathbb{P}_\Phi^\omega\{(Z_0, \dots, Z_n) \in C\} \\ &\quad = \int_{\mathcal{X}} \int_E \cdots \int_E \mathbf{1}_C\big((x, \omega), z_1, \dots, z_n\big) \mathbb{P}(z_{n-1}, dz_n) \\ &\qquad\qquad \cdots \mathbb{P}(z_1, dz_2) \mathbb{P}\big((x, \omega), dz_1\big) \varphi^\omega(dx). \end{aligned} \tag{5.13}$$

For each $(\Gamma, \mathbb{Q}) \in \mathcal{Q}_0$, fix a version $\mathbb{Q}_\Gamma^\omega$ of the conditional probability of $\mathbb{Q}_\Gamma$, given $\Theta_0 = \omega$.

5.14 LEMMA. *Assume* (A). *There exists a Borel set* Ω_l *of full* π*-measure such that, if* $\omega \in \Omega_l$, *then the following holds simultaneously for all* $(\Gamma, \mathbb{Q}) \in \mathcal{Q}_0$: *For all* n, *we have the derivatives*

$$f_n = \frac{d\mathbb{Q}_\Gamma^\omega}{d\mathbb{P}_\Phi^\omega}\bigg|_{\mathcal{F}(n, Z)}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\int \log f_n\, d\mathbb{Q}_\Gamma^\omega = H(\Gamma\otimes\mathbb{Q}\mid\Gamma\otimes\mathbb{P}). \tag{5.15}$$

If G is an open subset of $\mathcal{M}_1(E)$ containing Γ, then

$$\lim_{n\to\infty}\mathbb{Q}_\Gamma^\omega\{\widetilde{\mathbf{L}}_n\in G\}=1. \tag{5.16}$$

PROOF. Since $\mathcal{Q}_0$ is countable, we need to show that the lemma holds for a fixed $(\Gamma,\mathbb{Q})\in\mathcal{Q}_0$ for π-almost all ω. As observed above, $\mathbb{Q}(z,\cdot)\ll\mathbb{P}(z,\cdot)$ holds for Γ-almost all z. Without affecting the measure $\mathbb{Q}_\Gamma$, we may modify the kernel $\mathbb{Q}$ on a Γ-null set so that this holds for all z. Then we have a measurable function $g\colon E^2\to[0,\infty)$ so that $\mathbb{Q}(z_0,dz_1)=g(z_0,z_1)\mathbb{P}(z_0,dz_1)$, for all z_0 [for a proof, see, e.g., Nummelin (1984), Lemma 2.5], and

$$\infty > H(\Gamma\otimes\mathbb{Q}\mid\Gamma\otimes\mathbb{P})=\int\log g\, d\Gamma\otimes\mathbb{Q}.$$

By Lemma 5.10(ii) we have a Radon–Nikodym derivative $h=d\Gamma/d\Phi$, and

$$\infty > H(\Gamma\mid\Phi)=\int\log h\, d\Gamma.$$

Let $\gamma^\omega(dx)$ be the conditional distribution of Γ on $\mathcal{X}$, given $\omega\in\Omega$. Recalling similar notation introduced earlier for Φ, we may write $\Phi(dx,d\omega)=\varphi^\omega(dx)\pi(d\omega)$ and $\Gamma(dx,d\omega)=\gamma^\omega(dx)\pi(d\omega)$. It follows that $\gamma^\omega(dx)=h(x,\omega)\varphi^\omega(dx)$ for π-almost all ω. After throwing away a π-null set, we may assume that $\mathbb{Q}_\Gamma^\omega$ is given by

$$\begin{aligned}&\mathbb{Q}_\Gamma^\omega\{(Z_0,\dots,Z_n)\in C\}\\ &\quad=\int_{\mathcal{X}}\int_E\cdots\int_E \mathbf{1}_C\big((x,\omega),z_1,\dots,z_n\big)\mathbb{Q}(z_{n-1},dz_n)\\ &\qquad\cdots\mathbb{Q}(z_1,dz_2)\mathbb{Q}\big((x,\omega),dz_1\big)\gamma^\omega(dx).\end{aligned} \tag{5.17}$$

By comparing (5.13) and (5.17), it is evident that

$$f_n=h(Z_0)\prod_{j=0}^{n-2}g\big(Z_j,Z_{j+1}\big)$$

does the job. Apply Lemma 5.12 to conclude that

$$\frac{1}{n}\int\log f_n\,d\mathbb{Q}_\Gamma^\omega=\frac{1}{n}\int\log h(x,\omega)\gamma^\omega(dx)+\frac{1}{n}\sum_{j=0}^{n-2}\int\log g\big(Z_0,Z_1\big)d\mathbb{Q}_\Gamma^{T^j\omega},$$

π-almost surely. The first integral on the right-hand side is finite π-almost surely, so the term vanishes as $n \to \infty$. By the ergodic theorem, the limit of the second term is

$$\int \left[\int \log g(Z_0, Z_1) d\mathbb{Q}_\Gamma^\omega \right] \pi(d\omega) = \int \log g \, d\Gamma \otimes \mathbb{Q} = H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}).$$

This proves (5.15).

For a fixed G, (5.16) follows from ergodicity. To get it for all open sets simultaneously, let G vary over a countable base for the weak topology of $\mathcal{M}_1(E)$. □

PROOF OF PROPOSITION 5.7. By Assumption (A), for π-almost all ω and all $n \in \mathbb{N}$,

$$\frac{1}{n} \log \inf_{x \in \mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in G\} \geq -\frac{\log M(\omega)}{n} + \frac{1}{n} \log \mathbb{P}_\Phi^\omega\{\widetilde{\mathbf{L}}_n \in G\}. \tag{5.18}$$

Let $(\Gamma, \mathbb{Q}) \in \mathcal{Q}_0$ be such that $\Gamma \in G$. Since the set $\{\widetilde{\mathbf{L}}_n \in G\}$ is $\mathcal{F}(n+b, Z)$-measurable and $\mathbb{Q}_\Gamma^\omega\{f_{n+b} = 0\} = 0$, we may write

$$\frac{1}{n} \log \mathbb{P}_\Phi^\omega\{\widetilde{\mathbf{L}}_n \in G\} \geq \frac{1}{n} \log \int \mathbf{1}_G(\widetilde{\mathbf{L}}_n) f_{n+b}^{-1} \, d\mathbb{Q}_\Gamma^\omega,$$

and use Jensen's inequality to get

$$\begin{aligned} &\frac{1}{n} \log \mathbb{P}_\Phi^\omega\{\widetilde{\mathbf{L}}_n \in G\} \\ &\quad \geq \frac{1}{n} \log \mathbb{Q}_\Gamma^\omega\{\widetilde{\mathbf{L}}_n \in G\} \\ &\qquad - \left[\mathbb{Q}_\Gamma^\omega\{\widetilde{\mathbf{L}}_n \in G\}\right]^{-1} \frac{1}{n} \int \mathbf{1}_G(\widetilde{\mathbf{L}}_n) \log f_{n+b} \, d\mathbb{Q}_\Gamma^\omega. \end{aligned} \tag{5.19}$$

By the elementary inequality $x \log x \geq -1/e$,

$$-\frac{1}{n} \int \mathbf{1}_G(\widetilde{\mathbf{L}}_n) \log f_{n+b} \, d\mathbb{Q}_\Gamma^\omega \geq -\frac{1}{en} - \frac{1}{n} \int \log f_{n+b} \, d\mathbb{Q}_\Gamma^\omega,$$

so (5.18), (5.19) and Lemma 5.14 combine to give

$$\liminf_{n \to \infty} \frac{1}{n} \log \inf_{x \in \mathcal{X}} \mathbb{P}_{(x,\omega)}\{\widetilde{\mathbf{L}}_n \in G\} \geq -H(\Gamma \otimes \mathbb{Q} \mid \Gamma \otimes \mathbb{P}).$$

Since this holds for all $(\Gamma, \mathbb{Q}) \in \mathcal{Q}_0$ satisfying $\Gamma \in G$, we are done by Lemma 5.10. □

To complete the proof of Theorem 3.3, it remains to establish the properties of I_E. Lemma 2.3 and (3.1) show that $\mathbb{J}$ is convex and lower semicontinuous. The functional $\mathbb{J}$ has compact level sets by Lemma 5.9 (iii) and the fact that relative entropy has compact level sets. It is immediate from (3.2) that I_E inherits these properties. According to Deuschel and Stroock [(1989), Lemma 5.1.4], $\mathbb{J}(\Gamma) = 0$ if and only if $\Gamma\mathbb{P} = \Gamma$. By (3.2) and Theorem 2.8(i), $I_E(\Gamma) = 0$ if and only if $\Gamma = \Phi$.

5.2. *Position level with a multivariate $\mathcal{X}$-component.* Our strategy will be to apply Theorem 3.3 to a new setup where $\mathcal{X}$ is replaced $\mathcal{X}^d$. Since we formulated Theorem 3.3 for a general dynamical system instead of the canonical one, we need only "redirect" the map from Ω into $\mathcal{P}(\mathcal{X}^d)$.

Define a complete metric s_d on $\mathcal{X}^d$ by

$$s_d(\mathbf{x}, \mathbf{y}) = \sum_{k=0}^{d-1} s(x_k, y_k),$$

and let r_d be the corresponding Prohorov metric on $\mathcal{M}_1(\mathcal{X}^d)$. Recall that the metric for $\mathcal{P}(\mathcal{X})$ was defined in terms of a fixed countable base $\{C_k\}$ of relatively compact open sets for the topology of $\mathcal{X}$. For a multiindex $\mathbf{k} = (k_0, \dots, k_{d-1}) \in \mathbb{N}^d$, put $C(\mathbf{k}) = C_{k_0} \times \cdots \times C_{k_{d-1}}$. The set $C(\mathbf{k})$ is a relatively compact, open subset of $\mathcal{X}^d$, and the collection $\{C(\mathbf{k}): \mathbf{k} \in \mathbb{N}^d\}$ is a countable base for the topology of $\mathcal{X}^d$. Put $|\mathbf{k}| = k_0 + \cdots + k_{d-1}$. The corresponding metric for $\mathcal{P}(\mathcal{X}^d)$ is defined by

$$D_d(\mathbf{P}, \mathbf{Q}) = \sum_{\mathbf{k}} 2^{-|\mathbf{k}|} \sup_{\mathbf{x} \in C(\mathbf{k})} r_d\big(\mathbf{P}(\mathbf{x}), \mathbf{Q}(\mathbf{x})\big).$$

5.20 LEMMA. *Given $P \in \mathcal{P}(\mathcal{X})$, define $\mathbf{R} = \mathbf{R}(P) \in \mathcal{P}(\mathcal{X}^d)$ by*

$$\int_{\mathcal{X}^d} f(\mathbf{y}) \mathbf{R}(\mathbf{x}, d\mathbf{y}) = \int_{\mathcal{X}} f(x_1, \dots, x_{d-1}, y) P(x_{d-1}, dy),$$

for $f \in B(\mathcal{X}^d)$. Then the map $P \mapsto \mathbf{R}(P)$ from $(\mathcal{P}(\mathcal{X}), D)$ into $(\mathcal{P}(\mathcal{X}^d), D_d)$ satisfies $D_d(\mathbf{R}(P_1), \mathbf{R}(P_2)) \le D(P_1, P_2)$.

PROOF. That $\mathbf{R} = \mathbf{R}(P)$ is Feller continuous is obvious from $\mathbf{R}(\mathbf{x}) = \delta_{(x_1, \dots, x_{d-1})} \otimes P(x_{d-1})$, for $\mathbf{x} = (x_0, \dots, x_{d-1})$.

Now suppose $\mathbf{R}_1 = \mathbf{R}(P_1)$ and $\mathbf{R}_2 = \mathbf{R}(P_2)$, for some $P_1, P_2 \in \mathcal{P}(\mathcal{X})$. It suffices to show that $r_d(\mathbf{R}_1, \mathbf{R}_2) \le r(P_1(x_{d-1}), P_2(x_{d-1}))$, for all $\mathbf{x} = (x_0, \dots, x_{d-1}) \in \mathcal{X}^d$. Given $\mathbf{x} \in \mathcal{X}^d$, let $\varepsilon > r(P_1(x_{d-1}), P_2(x_{d-1}))$. For $A \in \mathcal{B}_{\mathcal{X}^d}$ we may write

$$\mathbf{R}_i(\mathbf{x}, A) = P_i(x_{d-1}, A[\mathbf{x}]), \qquad i = 1, 2, \tag{5.21}$$

where $A[\mathbf{x}] = \{y \in \mathcal{X}: (x_1, \dots, x_{d-1}, y) \in A\}$. By the choice of ε,

$$P_1(x_{d-1}, B) \le P_2(x_{d-1}, B^{\varepsilon}) + \varepsilon, \tag{5.22}$$

for all $B \in \mathcal{B}_{\mathcal{X}}$. Relations (5.21) and (5.22) and the fact that $(A[\mathbf{x}])^{\varepsilon} \subset (A^{\varepsilon})[\mathbf{x}]$ give $\mathbf{R}_1(\mathbf{x}, A) \le \mathbf{R}_2(\mathbf{x}, A^{\varepsilon}) + \varepsilon$, for all $A \in \mathcal{B}_{\mathcal{X}^d}$, hence $r_d(\mathbf{R}_1(\mathbf{x}), \mathbf{R}_2(\mathbf{x})) \le \varepsilon$. $\square$

Letting $P: \Omega \to \mathcal{P}(\mathcal{X})$ still denote the original map introduced in Section 2, define $\mathbf{P}: \Omega \to \mathcal{P}(\mathcal{X}^d)$ by $\mathbf{P} = \mathbf{R} \circ P \circ T^{d-1}$. To be more explicit, here is how $\mathbf{P}(\omega)$ acts on a function $f \in B(\mathcal{X}^d)$:

$$\int_{\mathcal{X}^d} f(\mathbf{y}) \mathbf{P}(\omega; \mathbf{x}, d\mathbf{y}) = \int_{\mathcal{X}} f(x_1, \dots, x_{d-1}, y) P(T^{d-1}\omega; x_{d-1}, dy).$$

5.23 PROPOSITION. *If Assumption* (A) *holds for* P *with* b *and* M, *it holds for* $\mathbf{P}$ *with* $b' = b + d - 1$ *and* $M' = M \circ T^{d-1}$.

PROOF. Use the definition of $\mathbf{P}$. □

The skew transition $\mathbb{P}^{(d)}$ on E_d is defined for $A \in \mathcal{B}_{\mathcal{X}^d}$ and $B \in \mathcal{B}_\Omega$ by

$$\mathbb{P}^{(d)}\big((\mathbf{x},\omega), A \times B\big) = \mathbf{P}(\omega;\ \mathbf{x}, A)\mathbf{1}_B(T_\omega).$$

Recall the probability measure Φ_d on E_d defined by

$$\Phi_d(C) = \mathbb{P}_\Phi\{(X_0,\dots,X_{d-1},\ \Theta_0) \in C\}.$$

5.24 PROPOSITION. *Under Assumption* (A), Φ_d *is the unique* $\mathbb{P}^{(d)}$*-invariant probability measure on* E_d *with* Ω*-marginal* π.

PROOF. The Ω-marginal of Φ_d is π by definition, and invariance follows by a straightforward computation. Uniqueness then follows from Proposition 5.23, by applying Theorem 2.8(i) to $\Phi_d, \mathbb{P}^{(d)}$ and E_d. □

5.25 DEFINITION. For $\Gamma \in \mathcal{M}_1(E_d)$, define $\Gamma' \in \mathcal{M}_1(E_d)$ by

$$\int f\,d\Gamma' = \iint f\Big((x_0,\dots,x_{d-2},y),\omega\Big)P\big(T^{d-2}\omega;\ x_{d-2}, dy\big)\Gamma(d\mathbf{x}, d\omega),$$

for $f \in B(E_d)$.

Note that the property of d-invariance defined by (3.4) is precisely what is needed for $\Gamma' = \Gamma\mathbb{P}^{(d)}$. The measure Γ' agrees with Γ on the σ-field $\mathcal{E}_{d-1}$ generated by the coordinates $((x_0,\dots,x_{d-2}),\omega)$, so

$$H(\Gamma \mid \Gamma') = \int H\big(\Gamma(\cdot \mid \mathcal{E}_{d-1}) \mid \Gamma'(\cdot \mid \mathcal{E}_{d-1})\big)\,d\Gamma. \tag{5.26}$$

A regular conditional probability of Γ', given $\mathcal{E}_{d-1}$, is given by

$$\Gamma'\big(\cdot \mid (x_0,\dots,x_{d-2}),\omega\big) = \delta_{(x_0,\dots,x_{d-2})} \otimes P\big(T^{d-2}\omega;\ x_{d-2},\cdot\big) \otimes \delta_\omega. \tag{5.27}$$

Let $\mathbb{J}^{(d)}$ denote the Donsker–Varadhan functional for the transition $\mathbb{P}^{(d)}$.

5.28 PROPOSITION. *Suppose* $d \geq 2$, *and let* $\Gamma \in \mathcal{M}_1(E_d)$. *Then*

$$\mathbb{J}^{(d)}(\Gamma) = \begin{cases} H\big(\Gamma \mid \Gamma\mathbb{P}^{(d)}\big), & \text{if } \Gamma \text{ is } d\text{-invariant}, \\ \infty, & \text{otherwise}. \end{cases}$$

PROOF. If Γ is not d-invariant, let $K > 0$ and find a $g \in C_b\big(E_{d-1}\big)$ such that $g \geq 0$ and

$$\int g((x_0,\dots,x_{d-2}),\omega)\Gamma(dx,d\omega) - \int g\big((x_1,\dots,x_{d-1}),T\omega\big)\Gamma(d\mathbf{x},d\omega) > K.$$

Put $u(\mathbf{x},\omega) = \exp[g((x_0,\dots,x_{d-2}),\omega)]$ and deduce that

$$\mathbb{J}^{(d)}(\Gamma) \geq \int \log \frac{u}{\mathbb{P}^{(d)}u}\, d\Gamma > K.$$

This shows that $\mathbb{J}^{(d)}(\Gamma) = \infty$.

Now suppose that Γ is d-invariant. We shall show that $\mathbb{J}^{(d)}(\Gamma) = H(\Gamma \mid \Gamma')$, which gives the conclusion. Let $u \in V_b(E)$ be arbitrary. By Jensen's inequality and the d-invariance of Γ,

$$\begin{aligned}\mathbb{J}^{(d)}(\Gamma) &\geq \int \log u\, d\Gamma - \log \int \mathbb{P}^{(d)}u\, d\Gamma \\ &= \int \log u\, d\Gamma - \log \int u\, d\Gamma',\end{aligned}$$

so $\mathbb{J}^{(d)}(\Gamma) \geq H(\Gamma \mid \Gamma')$. Conversely,

$$\begin{aligned}&\int \log u\, d\Gamma - \int \log \mathbb{P}^{(d)}u\, d\Gamma \\ &\quad = \int \left[\int \log u\, d\Gamma(\cdot \mid \mathcal{E}_{d-1}) - \log \int u\, d\Gamma'(\cdot \mid \mathcal{E}_{d-1})\right] d\Gamma \\ &\quad \leq \int H\big(\Gamma(\cdot \mid \mathcal{E}_{d-1}) \mid \Gamma'(\cdot \mid \mathcal{E}_{d-1})\big)\, d\Gamma \\ &\quad = H(\Gamma \mid \Gamma'),\end{aligned}$$

hence $\mathbb{J}^{(d)}(\Gamma) \leq H(\Gamma \mid \Gamma')$. $\square$

As an immediate corollary from this and definition (3.5) of $I^{(d)}$ we get

$$I^{(d)}(\Gamma) = \begin{cases} \mathbb{J}^{(d)}(\Gamma), & \text{if } \Gamma_\Omega = \pi, \\ \infty, & \text{otherwise.} \end{cases} \tag{5.29}$$

Comparing with (3.2), we see that $I_{E_d} \equiv I^{(d)}$, so Theorem 3.3 gives the properties of $I^{(d)}$ stated in the first part of Theorem 3.6.

Let $(\mathbf{X}_n, \Theta_n)$ denote the coordinate variables on $(E_d)^{\mathbb{Z}^+}$, and let

$$\mathbf{L}'_n = \frac{1}{n} \sum_{k=d-1}^{n+d-2} \delta_{(\mathbf{X}_k,\Theta_k)}.$$

By Theorem 3.3 and reasons explained in the first paragraph of subsection 5.1, $I^{(d)}$ governs the large deviations of $\mathbf{L}'_n$ under $\mathbb{P}^{(d)}_{(\mathbf{x},\omega)}$, uniformly in $\mathbf{x} \in \mathcal{X}^d$ and π-almost surely. To prove the large deviation principle of Theorem 3.6, it only remains to observe that the distribution of $\mathbf{L}'_n$ under $\mathbb{P}^{(d)}_{(\mathbf{x},T^{1-d}\omega)}$ is precisely the distribution of $\mathbf{M}^{(d)}_n$ under $\mathbb{P}_{(x,\omega)}$, if $\mathbf{x} = (x_0,\dots,x_{d-2},x)$.

5.3. *Process level for the skew chain.* The next lemma proves the first claim of Theorem 3.9.

5.30 LEMMA. *Under Assumption* (A), Φ_∞ *is* U*-invariant and* U*-ergodic.*

PROOF. The shift-invariance of $\mathbb{P}_\Phi$ translates into the U-invariance of Φ_∞. Suppose A is a U-invariant Borel subset of E_∞. Let $\widehat{A} = \{(\vec{X}_0, \Theta_0) \in A\}$ be its inverse image on $E^{\mathbb{Z}^+}$. Then $\widehat{A}$ is $\mathbb{P}_\Phi$-almost surely shift-invariant, so by the ergodicity of $\mathbb{P}_\Phi$, $\Phi_\infty(A) = \mathbb{P}_\Phi(\widehat{A}) = 0$ or 1. □

Given $\Gamma \in \mathcal{M}_1(E_\infty)$, let Γ_d be the marginal distribution of $((X_0, \dots, X_{d-1}), \Theta)$ on E_d. Define $I' : \mathcal{M}_1(E_\infty) \to [0, \infty]$ by

$$I'(\Gamma) = \sup_{d \geq 2} I^{(d)}(\Gamma_d) \vee I_E(\Gamma_1). \tag{5.31}$$

5.32 PROPOSITION. *The functional* I' *is lower semicontinuous, has compact level sets and* $I'(\Gamma) = 0$ *if and only if* $\Gamma = \Phi_\infty$. *Moreover,* I' *governs the large deviations of* $\mathbf{M}_n$ *under* $\mathbb{P}_{(x,\omega)}$, *uniformly over* $\mathcal{X}$ *and* π*-almost surely.*

PROOF. From the definition of I', $I'(\Gamma) = 0$ if and only if $I^{(d)}(\Gamma_d) = 0$ for all d. By Theorem 3.6, this is equivalent to $\Gamma_d = \Phi_d$ for all d, which is equivalent to $\Gamma = \Phi_\infty$. The rest of the proposition is an immediate consequence of the projective limit argument of large deviation theory. This allows us to deduce the process large deviation principle with rate I' from the succession of position results for $d = 1, 2, 3, \dots$. See de Acosta [(1990), Section 5], Dawson and Gärtner [(1987), Theorem 3.3], Deuschel and Stroock [(1989), Theorem 5.4.12] or Ellis and Wyner [(1989), Theorem 1.3]. □

With this proposition, what is needed to complete the proof of Theorem 3.9 is contained in the next lemma.

5.33 LEMMA. *Assume* (A). *For* $\Gamma \in \mathcal{M}_U(E_\infty)$, *the limit in the definition* (3.7) *of* $h(\Gamma \mid \Phi_\infty)$ *exists, and* $h(\Gamma \mid \Phi_\infty)$ *can be finite only if* $\Gamma_\Omega = \pi$. *The function* $h(\cdot \mid \Phi_\infty)$ *is an affine function on* $\mathcal{M}_U(E_\infty)$. *Furthermore,* $I^{(\infty)} = I'$ *on* $\mathcal{M}_1(E_\infty)$.

PROOF. Start by observing that a probability measure Γ on E_∞ is U-invariant if and only if each Γ_d is d-invariant in the sense of Definition 3.4. From this it follows that $I'(\Gamma) = \infty = I^{(\infty)}(\Gamma)$ whenever Γ is not U-invariant.

Now fix $\Gamma \in \mathcal{M}_U(E_\infty)$. Recall Definition 5.25 of Γ'_d. By U-invariance, $\Gamma'_d = \Gamma_d \mathbb{P}^{(d)}$. Proposition 5.28, (5.29), (5.31) and the fact that $\mathbb{J}^{(d)}(\Gamma_d)$ is increasing in d combine to give

$$I'(\Gamma) = \begin{cases} \lim\limits_{d \to \infty} H(\Gamma_d \mid \Gamma'_d), & \text{if } \Gamma_\Omega = \pi, \\ \infty, & \text{otherwise.} \end{cases} \tag{5.34}$$

By the definition of Φ_d and (5.27),

$$\Phi_d(X_{d-1} \in A \mid \mathcal{E}_{d-1}) = P(T^{d-2}\Theta_0; X_{d-2}, A) = \Gamma'_d(X_{d-1} \in A \mid \mathcal{E}_{d-1}),$$

so the conditional probabilities are equal: $\Phi_d(\cdot \mid \mathcal{E}_{d-1}) = \Gamma'_d(\cdot \mid \mathcal{E}_{d-1})$. Apply (5.3) $n-1$ times and the previous equality and (5.26), to obtain

$$\begin{aligned} H_{\mathcal{E}(n)}(\Gamma \mid \Phi_\infty) &= H(\Gamma_1 \mid \Phi) + \sum_{d=2}^{n} \int H\big(\Gamma_d(\cdot \mid \mathcal{E}_{d-1}) \mid \Phi_d(\cdot \mid \mathcal{E}_{d-1})\big)\, d\Gamma_d \\ &= H(\Gamma_1 \mid \Phi) + \sum_{d=2}^{n} \int H\big(\Gamma_d(\cdot \mid \mathcal{E}_{d-1}) \mid \Gamma'_d(\cdot \mid \mathcal{E}_{d-1})\big)\, d\Gamma_d \qquad (5.35) \\ &= H(\Gamma_1 \mid \Phi) + \sum_{d=2}^{n} H(\Gamma_d \mid \Gamma'_d). \end{aligned}$$

In case $H(\Gamma_1 \mid \Phi) = \infty$, both $h(\Gamma \mid \Phi_\infty)$ and $I'(\Gamma)$ are infinite, the former by the monotonicity of relative entropy and the latter by (5.31), (3.2) and Lemma 5.9(iii).

Suppose $H(\Gamma_1 \mid \Phi) < \infty$. Lemma 5.9(ii) implies that $\Gamma_\Omega = \pi$. Equations (5.34) and (5.35) give

$$\lim_{n\to\infty} \frac{1}{n} H_{\mathcal{E}(n)}(\Gamma \mid \Phi_\infty) = I'(\Gamma).$$

That $h(\cdot \mid \Phi_\infty)$ is affine on $\mathcal{M}_U(E_\infty)$ is proved as in Deuschel and Stroock [(1989), page 222]. □

Theorems 3.13 and 3.16 come from Theorems 3.6 and 3.9 via push-forwards: First define $F_d: E_d \to E^d$ and $F: E_\infty \to E^{\mathbb{Z}^+}$ by

$$F_d(\mathbf{x}, \omega) = \Big(\mathbf{x}, \big(\omega, T\omega, \dots, T^{d-1}\omega\big)\Big)$$

and

$$F(\xi, \omega) = \Big(\xi, \big(\omega, T\omega, T^2\omega, \dots\big)\Big).$$

The maps induced on the corresponding measure spaces then take $\mathbf{M}_n^{(d)}$ to $\mathbf{L}_n^{(d)}$ and $\mathbf{M}_n$ to $\mathbf{R}_n$, respectively. The straightforward but somewhat tedious details can be found in Seppäläinen (1991).

5.4. *The sample chains.* Theorems 3.19 and 3.30 follow immediately from Theorems 3.3 and 3.9 by push-forwards via the projections from E onto $\mathcal{X}$ and from E_∞ onto $\mathcal{X}^{\mathbb{Z}^+}$, respectively.

PROOF OF THEOREM 3.21. The expression

$$\log \sup_{x\in\mathcal{X}} \int \exp(S_n f)\, dP_x^\omega$$

is measurable as a function of ω, by Feller continuity. By a standard subadditivity argument,

$$c(f) = \lim_{n\to\infty} \frac{1}{n} \int \left[\log \sup_x \int \exp(S_n f)\, dP_x^\omega \right] \pi(d\omega).$$

The convexity of c follows from Hölder's inequality. If f, $g \in C_b(\mathcal{X})$ satisfy $g - \varepsilon \le f \le g + \varepsilon$, then $c(g) - \varepsilon \le c(f) \le c(g) + \varepsilon$, which gives Lipschitz continuity. The duality statement comes as in Deuschel and Stroock [(1989), Theorem 2.2.21], using the following uniform version of Varadhan's theorem [Deuschel and Stroock (1989), 2.1.18], valid for all ω for which Theorem 3.19 applies:

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \left| \frac{1}{n} \log \int \exp(S_n f)\, dP_x^\omega - \sup_{\mu \in \mathcal{M}_1(\mathcal{X})} \left\{ \int f\, d\mu - I_X(\mu) \right\} \right| = 0. \quad \square \tag{5.36}$$

PROOF OF THEOREM 3.25. By (3.18), it suffices to prove $J_\pi(\mu) = \mathbb{J}(\mu \otimes \pi)$. Let $u \in V_b(E)$ be arbitrary. Write u_ω for the element of $V_b(\mathcal{X})$ defined by $u_\omega(x) = u(x, \omega)$. Then $P(\omega; x, u_{T\omega}) = \mathbb{P}u(x, \omega)$ and, consequently,

$$J_\pi(\mu) \ge \iint \log \frac{u(x, \omega)}{\mathbb{P}u(x, \omega)} \mu(dx)\pi(d\omega).$$

We have $J_\pi(\mu) \ge \mathbb{J}(\mu \otimes \pi)$.

To get the converse, we use characterization (5.8) for the Donsker–Varadhan functional. Let $\mathbb{Q}$ be any Markov kernel on E with invariant measure $\mu \otimes \pi$ and such that

$$H(\mu \otimes \pi \otimes \mathbb{Q} \mid \mu \otimes \pi \otimes \mathbb{P}) = \iint H\big(\mathbb{Q}(x, \omega) \mid \mathbb{P}(x, \omega)\big) \mu(dx)\pi(d\omega) < \infty.$$

This forces $\mathbb{Q}$ to be of the form $\mathbb{Q}(x, \omega) = Q(\omega; x) \otimes \delta_{T\omega}$, where

$$Q(\omega; x, A) \equiv \mathbb{Q}\big((x, \omega), A \times \Omega\big), \qquad x \in \mathcal{X},\ A \in \mathcal{B}_{\mathcal{X}},$$

defines a Markov kernel $Q(\omega)$ on $\mathcal{X}$ which depends measurably on ω. An easy computation shows that μ is $Q(\omega)$-invariant for π-almost all ω. Hence again by (5.8), this time applied to each J_ω,

$$\begin{aligned} J_\pi(\mu) &\le \int H\big(\mu \otimes Q(\omega) \mid \mu \otimes P(\omega)\big) \pi(d\omega) \\ &= \iint H\big(Q(\omega; x) \mid P(\omega; x)\big) \mu(dx)\pi(d\omega) \\ &= \iint H\big(\mathbb{Q}(x, \omega) \mid \mathbb{P}(x, \omega)\big) \mu(dx)\pi(d\omega). \end{aligned}$$

Since $\mathbb{Q}$ was arbitrary, we have $J_\pi(\mu) \le \mathbb{J}(\mu \otimes \pi)$. $\square$

PROOF OF THEOREM 3.26. For $f \in C_b(\mathcal{X})$ define

$$\mathbf{c}(f) = \lim_{n\to\infty} \frac{1}{n} \log \sup_{x\in\mathcal{X}} \int \exp(S_n f)\, dP_x.$$

The limit exists by subadditivity. By Deuschel and Stroock [(1989), Lemma 5.1.36],

$$\mathbf{J}(\mu) = \sup\left\{ \int f\, d\mu - \mathbf{c}(f) \colon f \in C_b(\mathcal{X})\right\}.$$

By Theorem 3.21, it suffices to prove $c(f) \le \mathbf{c}(f)$, which follows from the equation for $c(f)$ in the proof of Theorem 3.21 and Jensen's inequality. □

PROOF OF THEOREM 3.33. Let $\mathcal{C}$ be the union of the $\mathcal{C}_n$. For $f \in \mathcal{C}$, let

$$S_n f = \sum_{k=0}^{n-1} f(\mathbf{X}_k)$$

and define

$$c(f) = \lim_{n\to\infty} \frac{1}{n} \int_\Omega \left[\sup_{x\in\mathcal{X}} \log \int \exp(S_n f)\, dP_x^\omega \right] \pi(d\omega). \tag{5.37}$$

Let Q be a fixed shift-invariant probability measure on $\mathcal{X}^{\mathbb{Z}^+}$. The analogue of (5.36) for process level and the fact that $\mathcal{C}$ generates the weak topology of $\mathcal{M}_1(\mathcal{X}^{\mathbb{Z}^+})$ imply that

$$h_X(Q) = \sup\left\{ \int f\, dQ - c(f) \colon f \in \mathcal{C}\right\}. \tag{5.38}$$

For $f \in \mathcal{C}_r$, Hölder's inequality applied to the right-hand side of (5.37) and the shift-invariance of π yield

$$c(f) \le \frac{1}{r} \int_\Omega \left[\sup_{x\in\mathcal{X}} \log \int e^{rf}\, dP_x^\omega \right] \pi(d\omega).$$

Apply this to $f \circ S^b$ and use Assumption (A) to get

$$c(f) \le \frac{1}{r+b} \int_\Omega \left[\log \iint \exp\left((r+b) f \circ S^b\right) dP_x^\omega\, \varphi^\omega(dx) \right] \pi(d\omega) + \frac{1}{r+b} \int \log M\, d\pi.$$

By the $\mathbb{P}$-invariance of Φ, $\varphi^\omega P(\omega) = \varphi^{T\omega}$. Apply this b times together with the shift-invariance of π, to arrive at

$$c(f) \le \frac{1}{r+b} \int_\Omega \left[\log \int e^{(r+b)f}\, dP_{\varphi^\omega}^\omega \right] \pi(d\omega) + \frac{1}{r+b} \int \log M\, d\pi.$$

Substitute this into (5.38) to get

$$h_X(Q) \ge \frac{1}{r+b}\left\{ \int (r+b)f\,dQ - \int_\Omega \left[\log \int e^{(r+b)f}\,dP^\omega_{\varphi^\omega}\right]\pi(d\omega)\right\} - \frac{1}{r+b}\int \log M\,d\pi.$$

Since f was arbitrary, we have

$$h_X(Q) \ge \frac{1}{r+b}K_r(Q) - \frac{1}{r+b}\int \log M\,d\pi,$$

and letting $r \uparrow \infty$ gives

$$h_X(Q) \ge \limsup_{r\to\infty} \frac{1}{r}K_r(Q).$$

The opposite inequality comes easily. Let $\gamma < h_X(Q)$ be arbitrary, and pick an integer r and $f \in \mathcal{C}_r$ so that

$$\int f\,dQ - c(f) > \gamma.$$

For large enough n,

$$\begin{aligned}&\frac{1}{n}\left\{ \int S_n f\,dQ - \int_\Omega \log \int \exp(S_n f)\,dP^\omega_{\varphi^\omega}\,\pi(d\omega)\right\}\\ &\qquad \ge \frac{1}{n}\left\{ \int S_n f\,dQ - \int_\Omega \sup_{x\in\mathcal{X}} \log \int \exp(S_n f)\,dP^\omega_x\,\pi(d\omega)\right\} > \gamma,\end{aligned}$$

where we also made use of Q's shift-invariance. However, $S_n f$ is an element of $\mathcal{C}_{n+r}$, hence $K_{n+r}(Q)/n > \gamma$, and letting $n \uparrow \infty$ gives

$$\liminf_{n\to\infty} \frac{1}{n}K_n(Q) \ge \gamma.$$

The proof is complete. □

For the final proof, we need to show that Proposition 2.1(iii) of Cogburn (1984) carries over. Recall that $\varphi^\omega(dx)$ is the conditional distribution of the $\mathcal{X}$ coordinate under Φ, given $\omega \in \Omega$. In the canonical setting, let $\mathcal{B}_\Omega^{-\infty,n}$ denote the σ-field on Ω generated by $(P_k: -\infty < k \le n)$. Then we have the following lemma.

5.39 LEMMA. *In the canonical setting, the map $\omega \mapsto \varphi^\omega$ is $\mathcal{B}_\Omega^{-\infty,-1}$-measurable.*

PROOF. Let $\mu^\omega(dx) = P^b(T^{-b}\omega; \hat{x}, dx)$, and let Π be the probability measure appearing in Lemma 4.5. The measure $\mu^\omega(dx)$ is a version of the conditional distribution of the $\mathcal{X}$ coordinate under Π, given ω. It is a function of the coordinates $P_{-b}, \dots, P_{-1}$, hence $\mathcal{B}_\Omega^{-\infty,-1}$-measurable. The formula

$$v\mathbb{P}^n(x,\omega) = \int_{\mathcal{X}} v(y, T^{-n}\omega)\frac{dP^n(T^{-n}\omega; y)}{d\mu^\omega}(x)\mu^{T^{-n}\omega}(dy)$$

shows that $v\mathbb{P}^n$ is $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_\Omega^{-\infty,-1}$-measurable, whenever v is a $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_\Omega^{-\infty,n-1}$-measurable function in $L^1(\Pi)$. Let $u = d\Phi/d\Pi$. Since $\Phi_\Omega = \Pi_\Omega = \pi$, it is also true that $\varphi^\omega(dx) = u(x,\omega)\mu^\omega(dx)$ for π-almost all ω. Thus it suffices to show that u is $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_\Omega^{-\infty,-1}$-measurable, and now Cogburn's proof applies word for word. □

PROOF OF THEOREM 3.37. Let $\Gamma \in \mathcal{M}_U(E_\infty)$ with marginal Q. Then $H_{\mathcal{F}(n,X)}(Q|\varphi_\infty) \le H_{\mathcal{E}(n)}(\Gamma \mid \Phi_\infty)$, for all n, which gives the first inequality. Now take $\Gamma = Q \otimes \pi$, and note that the conditional distribution of $\Gamma_d\mathbb{P}^{(d)}$ on $\mathcal{X}^d$, given $\Theta = \omega$, is $Q_{d-1} \otimes P(T^{d-2}\omega)$. Thus

$$H(\Gamma_d \mid \Gamma_d\mathbb{P}^{(d)}) = \int H\big(Q_d \mid Q_{d-1} \otimes P(\omega)\big)\pi(d\omega).$$

Note that the expressions are increasing in d. If both sides are infinite for some d, then $h(\Gamma \mid \Phi_\infty)$ and $h_\pi(Q)$ are both infinite, the former by (5.35) and the latter by its definition (3.36). Finiteness of the left-hand side implies that $\mathbb{J}(\Gamma_1)$ is finite, and hence so is $H(\Gamma_1 \mid \Phi)$ by Lemma 5.9(iii). Use again (5.35) and (3.36), and let $d \to \infty$ to get $h(\Gamma \mid \Phi_\infty) = h_\pi(Q)$. This and (3.28) complete the proof of the first statement of the theorem.

To justify $\varphi_\infty = \mathbf{P}_\varphi$ under Assumptions (A) and (B), write

$$\varphi_\infty(A) = \iint P_x^\omega(A)\varphi^\omega(dx)\pi(d\omega),$$

for a Borel subset A of $\mathcal{X}^{\mathbb{Z}^+}$. As functions of ω, $P_x^\omega(A)$ is $\mathcal{B}_\Omega^{0,\infty}$-measurable and by Lemma 5.39, φ^ω is $\mathcal{B}_\Omega^{-\infty,-1}$-measurable, so they are independent under π. We get

$$\varphi_\infty(A) = \iint P_x^\omega(A)\pi(d\omega)\varphi(dx) = \mathbf{P}_\varphi(A). \qquad \square$$

Acknowledgments. Part of this paper formed a portion of my thesis. It is dedicated to the memory of my advisor Steven Orey. Careful comments by an anonymous referee are also gratefully acknowledged.

REFERENCES

BAXTER, J. R., JAIN, N. C. and SEPPÄLÄINEN, T. O. (1993). Large deviations for nonstationary arrays and sequences. *Illinois J. Math.* **37** 302–328.

BOLTHAUSEN, E. (1987). Markov process large deviations in the τ-topology. *Stochastic Process. Appl.* **25** 95–108.

COGBURN, R. (1984). The ergodic theory of Markov chains in random environments. *Z. Wahrsch. Verw. Gebiete* **66** 109–128.

COGBURN, R. (1990). On direct convergence and periodicity for transition probabilities of Markov chains in random environments. *Ann. Probab.* **18** 642–654.

COGBURN, R. (1991). On the central limit theorem for Markov chains in random environments. *Ann. Probab.* **19** 587–604.

COMETS, F. (1989). Large deviation estimates for a conditional probability distribution. Applications to random interaction Gibbs measures. *Probab. Theory Related Fields* **80** 407–432.

DAWSON, D. A. and GÄRTNER, J. (1987). Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. *Stochastics* **20** 247–308.

DE ACOSTA, A. (1988). Large deviations for vector-valued functionals of a Markov chain: Lower bounds. *Ann. Probab.* **16** 925–960.

DE ACOSTA, A. (1990). Large deviations for empirical measures of Markov chains. *J. Theoret. Probab.* **3** 395–431.

DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large Deviations.* Academic, San Diego.

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

DONSKER, M. D. and VARADHAN, S. R. S. (1975b). Asymptotic evaluation of certain Markov process expectations for large time II. *Comm. Pure Appl. Math.* **28** 279–301.

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

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

ELLIS, R. S. (1988). Large deviations for the empirical measure of a Markov chain with an application to the multivariate empirical measure. *Ann. Probab.* **16** 1496–1508.

ELLIS, R. S. and WYNER, A. D. (1989). Uniform large deviation property of the empirical process of a Markov chain. *Ann. Probab.* **17** 1147–1151.

FOGUEL, S. R. (1969). *The Ergodic Theory of Markov Processes.* Van Nostrand, New York.

JAIN, N. C. (1990). Large deviation lower bounds for additive functionals of Markov processes. *Ann. Probab.* **18** 1071–1098.

MUNKRES, J. R. (1975). *Topology.* Prentice-Hall, Englewood Cliffs, NJ.

NEY P. and NUMMELIN, E. (1987). Markov additive processes II: Large deviations. *Ann. Probab.* **15** 593–609.

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge Univ. Press.

OREY, S. (1986). Large deviations in ergodic theory. In *Seminar on Stochastic Processes* (E. Çinlar, K.-L. Chung and R. Getoor, eds.) 195–249. Birkhäuser, Boston.

OREY, S. (1991). Markov chains with stochastically stationary transition probabilities. *Ann. Probab.* **19** 907–928.

OREY, S. and PELIKAN, S. (1988). Large deviation principles for stationary processes. *Ann. Probab.* **16** 1481–1495.

ROSENBLATT, M. (1971). *Markov Processes, Structure and Asymptotic Behavior.* Springer, Berlin.

SEPPÄLÄINEN, T. (1991). Large deviations for processes with stationarily random distributions. Ph.D. dissertation, Univ. Minnesota.

SEPPÄLÄINEN, T. (1993a). Large deviations for lattice systems I: Parameterized independent fields. *Probab. Theory Related Fields* **96** 241–260.

SEPPÄLÄINEN, T. (1993b). Large deviations for lattice systems II: Nonstationary independent fields. *Probab. Theory Related Fields* **97** 103–112.

STROOCK, D. W. (1984). *An Introduction to the Theory of Large Deviations.* Springer, New York.

VARADHAN, S. R. S. (1984). *Large Deviations and Applications.* SIAM, Philadelphia.

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
231 WEST 18TH AVENUE
COLUMBUS, OHIO 43210