The Annals of Probability
1988, Vol. 16, No. 4, 1481–1495

LARGE DEVIATION PRINCIPLES FOR STATIONARY PROCESSES

BY STEVEN OREY[1] AND STEPHAN PELIKAN[2]

University of Minnesota and University of Cincinnati

Large deviation theorems at the Donsker–Varadhan level-three type are established for certain classes of stationary processes. Proceeding from more general to more specific assumptions allows us to describe the rate function more explicitly.

Consider a discrete time stationary process (ω_n) with n varying over the integers Z and ω_n taking values in a complete separable metric space M. Then $\omega = (\omega_m) \in M^Z := \Omega$, and, using the product topology, Ω is again metrizable as a complete separable metric space. Let $\mathscr{M}(\Omega)$ or $\mathscr{M}$ denote the class of probability measures on the Borel sets of Ω. Let $\theta\colon \Omega \to \Omega$ be the shift $(\theta\omega)_n = \omega_{n+1}$. Let $\mathscr{M}_\theta$ be the class of $\mu \in \mathscr{M}$ such that $\mu = \mu \circ \theta^{-1}$, that is, the class of stationary measures determining stationary stochastic processes.

For $\omega \in \Omega$ let δ_ω be the element in $\mathscr{M}$ assigning mass 1 to $\{\omega\}$ and let ε_n be the $\mathscr{M}$-valued random variable whose value at ω is given by $\varepsilon_{n,\omega} = n^{-1}\sum_{k=0}^{n-1}\delta_{\theta^k\omega}$. Note that for a Borel subset Λ of Ω, $\varepsilon_{n,\omega}(\Lambda)$ represents n^{-1} times the number of k between 0 and $n-1$ such that $\theta^k\omega \in \Lambda$. Given $\mu \in \mathscr{M}_\theta$, the random variable ε_n has a distribution $\mu \circ \varepsilon_n^{-1}$, and our concern is with the validity of a large deviation principle and the identification of the corresponding rate function. Here we use the familiar terminology of [10]. Precise definitions are given in the body of the paper.

The pioneering work on this question is Donsker and Varadhan [2], where this problem was discussed (actually for continuous time) in case the underlying shift is Markovian with transition probabilities having good continuity and mixing properties. Extensions to some non-Markovian situations are included in [4].

In the present paper we always assume M to be compact. In Section 1 a certain mixing condition (RM) is introduced, and it is shown that this condition alone suffices for the validity of a uniform large deviation principle. The condition (RM) does not appear to be comparable with other familiar mixing conditions in ergodic theory. Section 2 briefly recapitulates some results from [4]. If (RM) is supplemented by a condition (CD) the rate function in the large deviation principle can be identified as a relative entropy as in [2]. In Section 3 a certain class of Gibbs measures is considered (the same class studied in Bowen [1]), it is shown that (RM) and (CD) hold, and an explicit identification for the

Received August 1986; revised August 1987.
[1]Supported in part by National Science Foundation Grant MCS-83-01080 and the Institute of Mathematics and Its Applications.
[2]Supported in part by National Science Foundation Grant DMS-86-01014.
AMS 1980 *subject classifications*. Primary 60F10; secondary 60G10.
Key words and phrases. Large deviations, stationary processes, Gibbs measures.

rate function is given. Finally, in Section 4 some examples and counterexamples are given, and connections with the work of Takahashi ([8] and [9]) are given.

In Bowen [1] the theory of Gibbs states is developed with the motivation of applying it to the symbolic dynamics of Anosov diffeomorphisms. We also have such applications in mind; these will be given in a separate paper.

For stationary Gaussian sequences the large deviation principle has been beautifully worked out in Donsker and Varadhan [3].

1. A large deviation principle. For every integer n let $(M_n, \mathscr{B}_n) = (M, \mathscr{B})$, where M is a compact separable metric space and $\mathscr{B}$ the corresponding Borel field. Let

$$(\Omega, \mathscr{F}) = \prod_{n=-\infty}^{\infty} (M_n, \mathscr{B}_n)$$

be the product space, endowed with the product topology. Let $\mathscr{M}(\Omega)$ denote the class of probability measures on $(\Omega, \mathscr{F})$, with the topology of weak convergence. Usually we write $\mathscr{M}$ for $\mathscr{M}(\Omega)$. Then Ω and $\mathscr{M}$ are also compact separable metric spaces. The shift θ and the class $\mathscr{M}_\theta$ have been defined previously. If $\omega \in \Omega$, ω^- denotes the one-sided sequence $(\dots, \omega_{-1}, \omega_0)$, let $\Omega^- = \{\omega^-: \omega \in \Omega\}$. For any $\mu \in \mathscr{M}$ the system $(\Omega, \mathscr{F}, \theta, \mu)$ is called a *shift*. Let $\mathscr{M}_\theta$ be the class of $\mu \in \mathscr{M}$ such that $\mu = \mu \circ \theta^{-1}$. For $-\infty < m < n < \infty$ let $\mathscr{F}_{m,n} = \prod_{i=m}^{n} \mathscr{B}_i$ and put $\mathscr{F}_m = \mathscr{F}_{m,m}$. For $\mu \in \mathscr{M}$, $\mu_{m,n}$ will denote the restriction of μ to $\mathscr{F}_{m,n}$ and $\mu_m = \mu_{m,m}$. We now introduce a metric on $\mathscr{M}$. Choose a sequence (φ_n) of continuous functions on Ω satisfying the following conditions: (i) the space of (φ_n) is dense in the continuous functions on Ω, (ii) $|\varphi_n(\omega)| \le 1$, $\omega \in \Omega$, $n = 1, 2, \dots$, (iii) for each n there exists a positive integer m such that φ_n is $\mathscr{F}_{-m,m}$-measurable. Now define

$$\operatorname{dist}(\nu, \mu) = \sum_{n=1}^{\infty} 2^{-n} \left| \int \varphi_n \, d\nu - \int \varphi_n \, d\mu \right|,$$

for ν and $\mu \in \mathscr{M}$.

For each $A \in \mathscr{F}$ the random variable $\mu(A | \mathscr{F}_{-\infty,0})$ is defined only up to μ-null sets. Under our assumptions there exist *regular conditional probabilities*, that is, the choice of random variables can be made so that $\mu(\cdot | \mathscr{F}_{-\infty,0})(\omega)$ is a probability measure on $(\Omega, \mathscr{F})$ for each ω. We shall denote a choice of such a regular conditional probability by $(\mu^*_{\omega^-})$ or simply μ^*, so that $\mu^*_\cdot(A)$ is a version of $\mu(A | \mathscr{F}_{-\infty,0})(\cdot)$. If $\mu \in \mathscr{M}_\theta$ we will always require that for every $\omega \in \Omega$, $B \in \mathscr{F}$ and every nonnegative integer m,

$$\mu^*_{\omega^-}\left(\theta^{-m} B | \mathscr{F}_{-\infty,m}\right)(\eta) = \mu^*_{(\theta^m \eta)^-}(B) \tag{1.1}$$

holds for $\mu^*_{\omega^-}$ a.e. $\eta \in \Omega$.

We will consider $\mu \in \mathscr{M}_\theta$ satisfying the following *ratio-mixing condition*.

(RM) There exists a nondecreasing function $m(n)$ such that $0 < m(n) < n$, $m(n)/n \to 0$ as $n \to \infty$, and

$$\lim_{n \to \infty} \frac{1}{n} \sup \left\{ \log \frac{\mu^*_{\eta^-}(A)}{\mu^*_{\omega^-}(A)} : \eta^- \in \Omega^-, \ \omega^- \in \Omega^-, \ A \in \mathscr{F}_{m(n),n} \right\} = 0.$$

Denote by ε_n the $\mathscr{M}$-valued random variable whose value at ω is

$$\varepsilon_{n,\omega} = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{\theta^k\omega}.$$

For a fixed $\mu \in \mathscr{M}_\theta$ and choice of $\mu^* = (\mu^*_{\omega^-})$, let Q_{n,ω^-} denote the distribution of ε_n under $\mu^*_{\omega^-}$, that is,

$$Q_{n,\omega^-}(\mathscr{A}) = \mu^*_{\omega^-}\left[\eta\colon \varepsilon_{n,\eta} \in \mathscr{A}\right],$$

for $\mathscr{A}$ a Borel subset of $\mathscr{M}$. Let $Q_{n,\mu}$ be the distribution of ε_n under μ. Also define

$$\mu_{\inf}(A) = \inf_{\omega^-\in\Omega^-} \mu^*_{\omega^-}(A), \qquad A \in \mathscr{F},$$

and

$$Q_n(\mathscr{A}) = \mu_{\inf}\left[\eta\colon \varepsilon_{n,\eta} \in \mathscr{A}\right] = \inf_{\omega\in\Omega^-} Q_{n,\omega^-}(\mathscr{A}).$$

The purpose of this section is to prove the following theorem.

Theorem 1.1. *Let $\mu \in \mathscr{M}_\theta$ with μ^* satisfying* (RM). *Then there exists a lower semicontinuous function $K\colon \mathscr{M} \to [0,\infty]$ such that*

(i) $\displaystyle\liminf_{n\to\infty} \frac{1}{n}\log \inf_{\omega^-\in\Omega^-} Q_{n,\omega^-}(\mathscr{A}) \geq -\inf\{K(\nu)\colon \nu \in \mathscr{A}\}$, *$\mathscr{A}$ open*,

(ii) $\displaystyle\limsup_{n\to\infty} \frac{1}{n}\log \sup_{\omega^-\in\Omega^-} Q_{n,\omega^-}(\mathscr{A}) \leq -\inf\{K(\nu)\colon \nu \in \mathscr{A}\}$, *$\mathscr{A}$ closed*.

Remark 1.1. The K in Theorem 1.1 is easily shown to be unique; see [4].

Remark 1.2. Our proof exploits the subadditivity of $Q_n(\mathscr{A})$ for suitable $\mathscr{A}$. It is quite similar to the proof given (by Theorem 6.2 of Stroock [7].)

We proceed by a sequence of lemmas. First we introduce the class Γ of all subsets $\mathscr{A}$ of $\mathscr{M}$ representable in the form

$$\mathscr{A} = \left\{\nu \in \mathscr{M}\colon \left|\int Y_i\, d\nu - c_i\right| < \varepsilon, i = 1,2,\dots,p\right\},$$

where p is some positive integer, c_i are real numbers, $\varepsilon < 0$ and Y_i are real-valued, continuous (hence bounded) random variables, and there exist nonnegative integers N_1 and N_2 so that Y_i is $\mathscr{F}_{-N_1,N_2}$-measurable for $1 \leq i \leq p$. Then Γ is a basis of convex open sets for the topology of $\mathscr{M}$. Denote by Γ_0 the subclass of Γ consisting of all $\mathscr{A} \in \Gamma$ for which $N_2 = 0$.

Lemma 1.2. *For nonnegative integers n and m and $\mathscr{A} \in \Gamma_0$,*

$$Q_{n+m}(\mathscr{A}) \geq Q_n(\mathscr{A})Q_m(\mathscr{A}).$$

PROOF. Observe that

$$\begin{aligned}\mu^*_{\omega^-}[\varepsilon_{n+m} \in \mathscr{A}] &= \mu^*_{\omega^-}\left[\left(\frac{m}{m+n}\varepsilon_m + \frac{n}{m+n}\varepsilon_n \circ \theta_m\right) \in \mathscr{A}\right] \\ &\geq \mu^*_{\omega^-}[\varepsilon_m \in \mathscr{A}, \varepsilon_n \circ \theta_m \in \mathscr{A}] \\ &= \int_{[\varepsilon_m \in \mathscr{A}]} \mu^*_{\omega^-}\left[\theta^{-m}[\varepsilon_n \in \mathscr{A}] | \mathscr{F}_{-\infty, m}\right](\eta)\mu^*_{\omega^-}(d\eta) \\ &= \int_{[\varepsilon_m \in \mathscr{A}]} \mu^*_{(\theta^m\eta)^-}[\varepsilon_n \in \mathscr{A}]\mu^*_{\omega^-}(d\eta) \\ &\geq Q_n(\mathscr{A})Q_m(\mathscr{A}),\end{aligned}$$

where the first inequality follows from the convexity of $\mathscr{A}$, the second equality holds because $[\varepsilon_m \in \mathscr{A}] \in \mathscr{F}_{-\infty, m}$ since $\mathscr{A} \in \Gamma_0$ and the third inequality uses (1.1). Taking the infinmum over Ω^- gives the lemma. □

LEMMA 1.3. *For $\mathscr{A} \in \Gamma_0$, $Q_n(\mathscr{A}) \equiv 0$ or $Q_n(\mathscr{A}) > 0$ for all sufficiently large n.*

PROOF. Let $\mathscr{A} \in \Gamma_0$ and assume $Q_m(\mathscr{A}) > 0$. The parameter ε enters in the definition of $\mathscr{A}$; let us write $\mathscr{A} = \mathscr{A}^\varepsilon$, and for $0 < \varepsilon'$ let $\mathscr{A}^{\varepsilon'}$ be defined like $\mathscr{A}^\varepsilon$ but with ε' in place of ε. Assume $\mathscr{A}^\varepsilon \neq \Omega$. For $\varepsilon' < \varepsilon$, $\mathscr{A}^{\varepsilon'} \subseteq \mathscr{A}^\varepsilon$ and indeed there exists $c > 0$ such that $\operatorname{dist}(\nu', \nu) > c$, whenever $\nu' \in \mathscr{A}^{\varepsilon'}$ and $\nu \notin \mathscr{A}^\varepsilon$. By taking $\varepsilon' < \varepsilon$ but sufficiently large we can ensure $Q_m(\mathscr{A}^{\varepsilon'}) > 0$. By Lemma 1.2 $Q_{km}(\mathscr{A}^{\varepsilon'}) > 0$ for $k = 1, 2, \dots$. For $n > m$ write $n = qm + r$, $0 < r < m$. Then $Q_{qm}(\mathscr{A}^{\varepsilon'}) > 0$. For n sufficiently large $\operatorname{dist}(\varepsilon_{qm}, \varepsilon_n) < c$, and so $[\varepsilon_{qm} \in \mathscr{A}^{\varepsilon'}] \subseteq [\varepsilon_n \subseteq \mathscr{A}^\varepsilon]$ and hence $Q_n(\mathscr{A}^\varepsilon) \geq Q_{qm}(\mathscr{A}^{\varepsilon'}) > 0$. □

LEMMA 1.4.

$$\lim_{n\to\infty} \frac{1}{n}\log Q_n(\mathscr{A}) = \sup \frac{1}{n}\log Q_n(\mathscr{A}) =: \Lambda(\mathscr{A}), \quad \mathscr{A} \in \Gamma_0.$$

PROOF. By Lemma 1.2 $-(1/n)\log Q_n(\mathscr{A})$ is a subadditive function of n for $\mathscr{A} \in \Gamma_0$; by Lemma 1.3 this function is strictly positive for all large n, or else identically infinite. This suffices. □

Consider $\mathscr{A} \in \Gamma_0$ and write $\mathscr{A} = \mathscr{A}^\varepsilon$ as before. Then

$$\lim_{n\to\infty} \frac{1}{n}\log Q_n(\mathscr{A}^\varepsilon) = \Lambda(\mathscr{A}^\varepsilon)$$

is an increasing function of ε, and hence it has at most denumerably many points of discontinuity.

Now let $\mathscr{A} \in \Gamma$, with ε, N_1, N_2 as in the definition of Γ. Let $\hat{\mathscr{A}} = \theta^{N_2}\mathscr{A}$ and note $\hat{\mathscr{A}} \in \Gamma^0$.

Define Γ^1 to be the subclass of Γ consisting of all $\mathscr{A} \in \Gamma$ such that the ε corresponding to $\mathscr{A}$ is a continuity point of the function $\Lambda(\hat{\mathscr{A}}^{\varepsilon})$. Then Γ^1 is a basis of convex sets for the topology of $\mathscr{M}$.

Lemma 1.5. *For $\mathscr{A} \in \Gamma^1$, $\lim_{n \to \infty}(1/n)\log Q_n(\mathscr{A}) = \Lambda(\mathscr{A})$ exists.*

Proof. Let $\mathscr{A} \in \Gamma^1$, then $\hat{\mathscr{A}} \in \Gamma^0$ and if ε corresponds to $\mathscr{A}$, it is a continuity point of $\Lambda(\hat{\mathscr{A}}^{\varepsilon})$. For $\varepsilon' < \varepsilon < \varepsilon''$ and n sufficiently large

$$\left[\varepsilon_n \in \hat{\mathscr{A}}^{\varepsilon'}\right] \subseteq \left[\varepsilon_n \in \mathscr{A}^{\varepsilon}\right] \subseteq \left[\varepsilon_n \in \hat{\mathscr{A}}^{\varepsilon''}\right].$$

Applying Q_n, taking logarithms and dividing by n and then letting $n \to \infty$, the first and last terms go to $\Lambda(\hat{\mathscr{A}}^{\varepsilon'})$ and $\Lambda(\hat{\mathscr{A}}^{\varepsilon''})$, respectively. Since ε is a continuity point of $\Lambda(\hat{\mathscr{A}}^{\varepsilon})$ the result follows. □

Now we define

$$K(\nu) = -\inf\{\Lambda(\mathscr{A}) \colon \nu \in \mathscr{A}, \mathscr{A} \in \Gamma^1\}.$$

It is easily verified that K is lower semicontinuous and convex, and if $\mathscr{A}$ is closed and $\mathscr{A}^{(\delta)}$ is a δ-neighborhood of $\mathscr{A}$, then

$$\liminf_{\delta \downarrow 0}\{K(\nu) \colon \nu \in \mathscr{A}^{(\delta)}\} = \inf\{K(\nu) \colon \nu \in \mathscr{A}\}. \tag{1.2}$$

(Complete details for the analogous results in [7] are given in Section 6 of that monograph.)

Proof of Theorem 1.1. To prove (i), let $\nu \in \mathscr{A}$, where $\mathscr{A}$ is open. There exists $\mathscr{A}^1 \in \Gamma^1$ such that $\nu \in \mathscr{A}^1$ and $\mathscr{A}^1 \subseteq \mathscr{A}$. Then $Q_n(\mathscr{A}) \geq Q_n(\mathscr{A}^1)$ and so

$$\liminf_{n \to \infty} \frac{1}{n} \log Q_n(\mathscr{A}) \geq \Lambda(\mathscr{A}^1) \geq -K(\nu).$$

Taking the supremum over all $\nu \in \mathscr{A}$ in the last relation gives (i).

For the proof of (ii) choose $\mathscr{A}$ to be closed (hence compact). Let $c > -\inf\{K(\nu) \colon \nu \in \mathscr{A}\}$. For $\alpha > 0$, choose for each $\nu \in \mathscr{A}$ a neighborhood $\mathscr{N}$ of ν of diameter less than α with $\mathscr{N} \in \Gamma^1$ and $\Lambda(\mathscr{N}) < c$. Recall the constants N_1 and N_2 entering the definition of $\mathscr{N}$. With $m(n)$ the function introduced in (RM), define

$$k(n) = m(N_2 + 2n) + N_1.$$

Then $k(n) < n$ for $n > n_0$, say. This guarantees

$$-N_1 + k(n) > m(k(n) + N_2 + n). \tag{1.3}$$

There exist neighborhoods $\mathscr{N}'$ and $\mathscr{N}''$ of ν and a positive integer n' such that $\mathscr{N}' \in \Gamma^1$, $\mathscr{N}'' \subseteq \mathscr{N}' \subseteq \mathscr{N}$ and for $n \geq n'$,

$$[\varepsilon_n \in \mathscr{N}] \supseteq \theta^{k(n)}[\varepsilon_n \in \mathscr{N}'] \supseteq [\varepsilon_n \in \mathscr{N}''].$$

The event in the middle is $\mathscr{F}_{-N_1 + k(n),\, k(n) + N_2 + n}$-measurable. Now choose a finite set $\nu_1, \dots, \nu_q$ such that the corresponding $\mathscr{N}_1'', \dots, \mathscr{N}_q''$ cover $\mathscr{A}$. For each i,

$1 \le i \le q$, there will be a corresponding $\mathcal{N}_i$ depending on integers $N_1^{(i)}, N_2^{(i)}$. Now

$$(1.4)\qquad \begin{aligned} \mu^*_{\omega^-}[\varepsilon_n \in \mathscr{A}] &\le \sum_{i=1}^{q} \mu^*_{\omega^-}[\varepsilon_n \in \mathcal{N}_i''] \\ &\le \sum_{i=1}^{q} \mu^*_{\omega^-}[\theta^{k(n)}[\varepsilon_n \in \mathcal{N}_i']]. \end{aligned}$$

Let $k^i(n)$ be defined like $k(n)$ but with $N_1^{(i)}$ and $N_2^{(i)}$ for N_1 and N_2. Put

$$\delta_i(n) = \sup_{\omega^- \in \Omega^-} \left(\log \mu^*_{\omega^-}[\theta^{k_i(n)}[\varepsilon_n \in \mathcal{N}_i']]\right) - \log \mu_{\inf}[\theta^{k_i(n)}[\varepsilon_n \in \mathcal{N}_i']]$$

and observe that (1.3) allows us to apply (RM) to conclude that $\delta_i(n)/n \to 0$ as $n \to \infty$. Then from (1.4)

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\omega^- \in \Omega^-} \mu^*_{\omega^-}[\varepsilon_n \in \mathscr{A}] \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \sum_{i=1}^{q} \sup_{\omega^- \in \Omega^-} \mu^*_{\omega^-}[\theta^{k(n)}[\varepsilon_n \in \mathcal{N}_i']] \\ &= \max_{1\le i\le q} \limsup_{n\to\infty} \frac{1}{n} \log \sup_{\omega^- \in \Omega^-} \mu^*_{\omega^-}[\theta^{k(n)}[\varepsilon_n \in \mathcal{N}_i']] \\ &= \max_{1\le i\le q} \limsup_{n\to\infty} \frac{1}{n} \left(\log \mu_{\inf}[\theta^{k(n)}[\varepsilon_n \in \mathcal{N}_i']] + \delta_i(n)\right) \\ &\le \max_{1\le i\le q} \limsup_{n\to\infty} \frac{1}{n} (\log \mu_{\inf}[\varepsilon_n \in \mathcal{N}_i] + \delta_i(n)) \\ &= \max_{1\le i\le q} \Lambda(\mathcal{N}_i) \le c \end{aligned}$$

and since c was an arbitrary number greater than $-\inf\{K(\nu): \nu \in \mathscr{A}\}$ (ii) follows. □

2. Entropy. In their work [2] Donsker and Varadhan not only prove the existence of a rate function K, but they identify it as a relative entropy. As was shown in [4] their proof can be adapted to more general situations. In addition to the hypothesis (RM) of Section 1, one needs an assumption of *continuous dependence*:

(CD) For every real-valued continuous random variable Y which is $\mathscr{F}_{-\infty,1}$-measurable, $\omega^- \to \mu^*_{\omega^-}(Y)$ is a continuous function.

Let $\mu \in \mathscr{M}_\theta$, $\nu \in \mathscr{M}_\theta$ and let $(\mu^*_{\omega^-}), (\nu_{\omega^-})$ denote regular conditional probabilities given $\mathscr{F}_{-\infty,0}$ as in Section 1; by $\mu^*_{\omega^-}|_1$ and $\nu_{\omega^-}|_1$ we denote the restrictions to $\mathscr{F}_1$. The entropy $H(\nu; \mu^*)$ will now be defined. Set $H(\nu; \mu^*) = \infty$ if $\nu \notin \mathscr{M}_\theta$ or if $\nu_{\omega^-}|_1 \ll \mu_{\omega^-}|_1$ fails on a set of ω^- having positive ν-measure. In the remaining cases

$$H(\nu;\mu^*) = \int_{\Omega^-}\left[\int_M \log\frac{d\nu_{\omega^-}|_1}{d\mu^*_{\omega^-}|_1}(y)\nu_{\omega^-}|_1(dy)\right]\nu(d\omega^-).$$

Note that $H(\nu;\mu^*)$ depends on μ^* and not just μ, because one is integrating with respect to ν and μ-null sets will not usually be ν-null sets.

Throughout this paper we take M to be compact.

The next result, Theorem 2.1, is included in [4]. We include a brief proof, taking this opportunity to make explicit some details not given in [4].

THEOREM 2.1. *Let* $\mu \in \mathscr{M}_\theta$ *with* μ^* *satisfying* (RM) *and* (CD). *Then Theorem* 1.1 *holds with* $K(\nu) = H(\nu;\mu^*)$.

PROOF. It will be shown how to adapt the ideas of Donsker and Varadhan [2].

Actually in [2] it is not the random variable ε_n but a closely related random variable ε'_n which is studied. For each positive integer n define $\pi_n: \Omega \to \Omega$ by $(\pi_n\omega)_i = \omega_i$, $0 \le i < n$, and $(\pi_n\omega)_{j+n} = (\pi_n\omega)_j$, for all j. Then $\varepsilon'_{n,\omega} := \varepsilon_{n,\pi_n(\omega)}$. Note $\varepsilon'_{n,\omega} \in \mathscr{M}_\theta$. As $n \to \infty$, $\operatorname{dist}(\varepsilon_{n,\omega}, \varepsilon'_{n,\omega}) \to 0$ uniformly in ω, and the distributions of the random variables (ε_n) have the same rate function as those of the random variables (ε'_n). Hence K is infinite off $\mathscr{M}_\theta$.

In [2] the underlying measure is assumed to be Markovian. We can easily reduce our situation to that one by keeping track of the past. Formally consider the space $\hat{\Omega}$ of all bilateral sequences (η_n) with $\eta_n = (\dots, \eta_{n,-1}, \eta_{n,0}) \in \Omega^-$. For $\omega = (\omega_n) \in \Omega$ let $\hat{\omega} = (\hat{\omega}_n) \in \hat{\Omega}$ be defined by $\hat{\omega}_n = (\dots, \omega_{n-1}, \omega_n)$. This mapping takes μ on Ω onto $\hat{\mu}$ on $\hat{\Omega}$ and $\hat{\mu}$ is Markovian, corresponding to a Markov chain with values in Ω^-. The measures $\mu^*_{\omega^-}$ now correspond to the measures for the Markov chain started at time 0 at position ω^-. The condition (CD) is just the Feller property for the Markov chain. By the results of [2] we have a uniform large deviation principle with rate function $K(\hat{\nu}) = H(\hat{\nu};\mu^*)$ if $\hat{\nu}$ is stationary, and $K(\hat{\nu}) = \infty$ otherwise. Observe that actually only the closed subspace $\hat{\Omega}_1$ of $\hat{\Omega}$ consisting of those sequences (η_n) such that $\eta_{n,k} = \eta_{n+1,k-1}$ for every integer n and nonpositive integer k plays a role. In particular, $K(\hat{\nu}) < \infty$ implies $\hat{\nu}(\hat{\Omega}_1) = 1$. Finally, map back to the original space by $\Phi: \hat{\Omega} \to \Omega$ defined by $\Phi(\eta) = \omega$, where $\omega_n = \eta_{n,0}$. This map is continuous and Φ restricted to $\hat{\Omega}_1$ is a homeomorphism. To see that one obtains the upper bound (ii) of Theorem 1.1 with $H(\nu;\mu^*)$ in place of K, one only needs to note that for $\nu \in \mathscr{M}_\theta$ there is a unique stationary $\hat{\nu}$ on $\hat{\Omega}_1$ with $\nu = \hat{\nu} \circ \Phi^{-1}$ and from the definition of entropy $H(\nu;\mu^*) = H(\hat{\nu};\mu^*)$.

Finally, we must argue that the lower bound (i) of Theorem 1.1 holds with $H(\nu;\mu^*)$ in place of K. Our Markov process (ω^*) will never satisfy the assumptions of [2] (there is not even a reference measure in our case). However, going back to the original shift on Ω, the condition (RM) is exactly what is needed to imitate [2]. We sketch this briefly.

Start first with $\nu \in \mathscr{M}_\theta$ which is ergodic and satisfies $H(\nu;\mu^*) < \infty$. Using the properties of H, one obtains as in [2] that $(\nu_{\omega^-})_{0,n} \ll (\mu_{\omega^-})_{0,n}$ with corresponding

Radon–Nikodym derivative $\psi_n(\omega)$ $[\psi_n(\omega) = \psi_n(\omega^-; \omega_1, \dots, \omega_n)]$. Then $\psi_{n+m} = \psi_n \cdot \psi_m \circ \theta^n$, ν a.e., so that the ergodic theorem immediately implies the following "conditional Shannon–McMillan theorem":

$$\lim_{n\to\infty} \frac{1}{n} \log \psi_n = H(\nu; \mu^*), \quad \nu \text{ a.e.} \tag{2.1}$$

Now for $\alpha > 0$,

$$\mu^*_{\omega^-}[\varepsilon'_n \in \mathscr{N}] \geq \int_{[\varepsilon'_n \in \mathscr{N}]} e^{-\log \psi_n} \, d\nu_{\omega^-}$$
$$\geq e^{-n(H(\nu; \mu^*)+\alpha)} \nu_{\omega^-}\left[\varepsilon'_n \in \mathscr{N}, \frac{1}{n}\log \psi_n \leq H(\nu; \mu^*) + \alpha\right].$$

Let $\mathscr{N}$ be a neighborhood of ν. Then as $n \to \infty$ the second factor in the last member tends to 1, ν a.e., so that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu^*_{\omega^-}[\varepsilon'_n \in \mathscr{N}] \geq -H(\nu; \mu^*), \quad \nu \text{ a.e.} \tag{2.2}$$

To obtain Theorem 1.1(i), we need (2.2) to hold for all ω^-, and indeed the inequality is to hold uniformly in ω^-. Let $m(n)$ be as in condition (RM) of Section 1. Since $m(n)/n \to 0$, there exist neighborhoods $\mathscr{N}'$ and $\mathscr{N}''$ of ν and a positive integer n' such that $\mathscr{N} \supseteq \mathscr{N}' \supseteq \mathscr{N}''$ and

$$[\varepsilon'_n \in \mathscr{N}] \supseteq \theta^{-m(n)}[\varepsilon'_n \in \mathscr{N}'] \supseteq [\varepsilon'_n \in \mathscr{N}''], \qquad n > n'. \tag{2.3}$$

According to (2.2) there exists an $\omega^- \in \Omega^-$ and for every $\alpha > 0$ an n_α such that

$$\frac{1}{n}\log \mu^*_{\omega^-}[\varepsilon'_n \in \mathscr{N}''] \geq -\left(H(\nu; \mu^*) + \frac{\alpha}{2}\right), \qquad n > n_\alpha.$$

Now it follows from (2.3) and condition (RM) that there exists n'_α such that

$$\frac{1}{n}\log \mu^*_{\eta^-}[\varepsilon'_n \in \mathscr{N}] \geq -(H(\nu; \mu^*) + \alpha), \qquad n \geq n'_\alpha, \ \eta^- \in \Omega^-.$$

Since $\text{dist}(\varepsilon'_{n,\omega}, \varepsilon_{n,\omega}) \to 0$ uniformly, the corresponding relation with ε_n in place of $\varepsilon_n{}'$ is also valid.

It remains to consider the case that $\nu \in \mathscr{M}_\theta$ is not ergodic. This can be handled exactly as in [2]. □

3. Gibbs measures. Our compact space will now be a finite set $M = \{0, 1, \dots, N-1\}$. For any finite sequence $(x_0, x_1, \dots, x_{n-1})$ let

$$[x_0, x_1, \dots, x_{n-1}] = \{\omega \in \Omega(M) \colon \omega_i = x_i, 0 \leq i < n\}.$$

Also set $B_i = \{\omega \in \Omega(M) \colon \omega_1 = i\}$.

Given $\mu \in \mathscr{M}_\theta$, the condition (CD) of the previous section reduces to

$$\omega^- \to \mu^*_{\omega^-}(B_i) \text{ is a continuous function}, \qquad 0 \leq i < N. \tag{3.1}$$

For given a conditional probability distribution $(\mu^*_{\omega^-})$ defined only for $\{B_i \colon 0 \leq i < N\}$ one naturally extends it to $\mathscr{F}$. For example, if $\omega^- =$

$(\dots, \omega_{-1}, \omega_0)$, $B = [i, j, k]$, $\mu^*_{\omega^-}(B)$ is defined as follows: Let $\omega^- j = (\dots, \omega_{-1}, \omega_0, j)$ and set

$$\mu^*_{\omega^-}(B) = \delta_{\omega_0, i}\mu^*_{\omega^-}(B_j)\mu^*_{\omega^- j}(B_k).$$

If (3.1) holds μ^* will satisfy (CD).

Instead of $\Omega(M)$ we will also consider a subshift Ω_A determined by an $N \times N$ matrix A of 0's and 1's. That is, $\Omega_A = \{\omega \in \Omega: A_{\omega_i \omega_{i+1}} = 1 \text{ for all } i\}$. All concepts can be relativized to Ω_A and this will be indicated by an A in the notation, e.g., $\mathscr{F}^A = \{B \cap \Omega_A: B \in \mathscr{F}\}$, $\mathscr{M}^A$ is the class of probability measures on $(\Omega_A, \mathscr{F}^A)$ and $\mathscr{M}^A_\theta$ is the class of θ-invariant members of $\mathscr{M}^A$. Any $\mu \in \mathscr{M}^A_\theta$ may be extended in exactly one way to a $\mu \in \mathscr{M}_\theta$; evidently $\mu(B) = 0$ for $B \in \mathscr{F} \setminus \mathscr{F}_A$. This will allow us to start with $\mu \in \mathscr{M}^A_\theta$ and apply the results of the previous section. In working with Ω_A we will assume that there exists a positive integer m such that

$$(A^m)_{ij} > 0, \qquad 0 \le i < N, 0 \le j < N. \tag{3.2}$$

A finite or infinite sequence (x_i) of nonnegative integers less than N is *A-admissible* if $A_{x_i x_{i+1}} = 1$ for every pair of consecutive elements x_i and x_{i+1} of x.

Proposition 3.1. *Assume $\mu \in \mathscr{M}^A_\theta$ satisfies* (3.1) *with ω^- varying over Ω^-_A and also*

$$\inf\{\mu^*_{\omega^-}(B_i): \omega^- i \text{ is } A\text{-admissible}\} > 0. \tag{3.3}$$

Then (RM) *holds with $m(n) = m$. Considering μ as an element of $\mathscr{M}_\theta$, one can define $(\mu^*_{\omega^-})$ so that* (CD) *and* (RM) *continue to hold for the extended family.*

Proof. If x^- and y^- belong to Ω^-_A and $x^- i$ and $y^- i$ also belong to Ω^-_A we can write

$$\mu^*_{x^-}(B_i) = \mu^*_{y^-}(B_i)(1 + \alpha(x^-, y^-, i))$$

and define

$$\alpha_s = \sup\{\alpha(x^-, y^-, i): x_k = y_k, -s < k \le 0, 0 \le i < N\}.$$

Our assumptions imply

$$\lim_{s \to \infty} \alpha_s = 0. \tag{3.4}$$

In verifying (RM) it will suffice to consider the supremum over all sets $B \in \mathscr{F}^A_{m,n}$ having the form $B = \theta^{-m}[i_m, i_{m+1}, \dots, i_n]$. Then

$$\begin{aligned}\mu^*_{\omega^-}(B) = \sum_k \mu^*_{\omega^-}(B_{k_1})\mu^*_{\omega^- k_1}(B_{k_2}) \cdots \mu^*_{\omega^- k_1 \cdots k_{m-1}}(B_{i_m}) \cdots \\ \mu^*_{\omega^- k_1 \cdots k_{m-1} i_m \cdots i_{n-1}}(B_{i_n}),\end{aligned} \tag{3.5}$$

where the sum extends over all $k = (k_1, \dots, k_{m-1})$ such that $(\omega_0, k_1, \dots, k_{m-1}, i_m)$ is admissible; by our assumption the sum contains at least one term, and at most N^{m-1} terms and each factor in the sum is bounded below by a positive constant, by (3.1), and bounded above by a constant times $\prod_{t=0}^{n-m}(1+\alpha_t)$, and by (3.4)

$$\lim_{n\to\infty} n^{-1}\log \prod_{t=0}^{n-m}(1+\alpha_t) = 0.$$

The final assertion of the proposition is easily seen to be true. □

Following the exposition of Bowen [1], a certain class of $\mu \in \mathscr{M}_\theta^A$ will be introduced and called *Gibbs measures*. (According to traditional terminology this class ought to be called "Gibbs measures with translation invariant exponentially decreasing interactions"; see Ruelle [5], Section 5.18.)

For $\beta \in (0,1)$ define the metric $d_\beta(\omega,\eta)$ on Ω_A by $d_\beta(\omega,\eta) = \beta^n$, where n is the least nonnegative integer such that $\omega_n \neq \eta_n$ or $\omega_{-n} \neq \eta_{-n}$. Let H_A denote the class of real-valued functions ϕ on Ω_A which are Hölder continuous. H_A does not depend on β (though changing β may change the Hölder exponent). Note $\phi \in H_A$ if and only if there exists a positive constant b and $\gamma \in (0,1)$ such that

$$\sup\{|\phi(x)-\phi(y)|: x \in \Omega_A, y \in \Omega_A, x_i = y_i \text{ for } |i| < n\} < b\gamma^n,$$
$$n = 0,1,\dots .$$

We rely on the following theorem: *For $\phi \in H_A$ there exist a unique $\mu \in \mathscr{M}_\theta^A$ and unique real number P for which one can find positive constants c_1 and c_2 such that*

$$c_1 \le \frac{\mu[\omega_0,\omega_1,\dots,\omega_{n-1}]}{\exp\{-Pn + \sum_{k=0}^{n-1}\phi(\theta^k\omega)\}} \le c_2, \tag{3.6}$$

for every $\omega \in \Omega_A$ and $n > 0$. For a proof of the theorem see [1], Theorems 1.2 and 1.22. The measure μ is called the *Gibbs measure* corresponding to ϕ. (The constant P is equal to the pressure of ϕ; see [1], Theorem 1.22. If ϕ and ϕ' belong to H_A they determine the same Gibbs measure if and only if there exists a constant c and a $\psi \in H_A$ such that $\phi - \phi' = c + \psi\circ\theta - \psi$; see Ruelle [5]. Theorem 5.21.) For $\phi \in H_A$ there exists $\phi' \in H_A$ such that $\phi'(x) = \phi'(y)$ for all $x \in \Omega_A$, $y \in \Omega_A$, with $x_i = y_i$ for $i > 0$, and ϕ' determines the same Gibbs measure as ϕ; see [1], Lemma 1.6.

Given $\nu \in \mathscr{M}$, the notations (ν,ψ) and $\int\psi\, d\nu$ will be used with the same significance. If $\nu \in \mathscr{M}_\theta$, h_ν will denote the Kolmogorov–Sinai entropy of ν with respect to the shift θ. The principal result of the section can now be stated.

THEOREM 3.2. *Let $\phi \in H_A$, μ the corresponding Gibbs measure extended to Ω, P the pressure, as in* (3.6). *Then there exists μ^* satisfying* (CD) *and* (RM). *The large deviation principle as given in Theorem* 1.1 *holds and*

$$K(\nu) = H(\nu;\mu^*) = -h_\nu - (\nu,\phi) + P, \qquad \nu \in \mathscr{M}_\theta, \tag{3.7}$$

where $(\nu,\phi) := \infty$ if $\nu \in \mathscr{M}_\theta \setminus \mathscr{M}_\theta^A$.

The conditional probabilities $\mu^*_{x^-}(\theta^{-1}[i])$, with $x^- \in \Omega^-$ and $0 \le i < N$, will be obtained as limits. Assume μ is a Gibbs measure. Let

$$\mu^{(n)}_{x^-}(i) = \frac{\mu[x_{-n}, \dots, x_0, i]}{\mu[x_{-n}, \dots, x_0]}, \tag{3.8}$$

which is well defined if $(x_{-n}, \dots, x_0, i)$ is A-admissible; otherwise make the following convention: $\mu^{(n)}_{x^-}(i) = 0$ if $A_{x_0 i} = 0$, and if $A_{x_0 i} = 1$, $\mu^{(n)}_{x^-}(i) = \mu^{(s)}_{x^-}(i)$, where s is the largest integer such that $(x_{-s}, \dots, x_0, i)$ is A-admissible.

Statements similar to the following lemma can be found in the literature (see Remark 4.5), but we have found no place where a proof is given.

Lemma 3.3. *For a Gibbs measure μ, $\mu^{(n)}_{x^-}(i)$ converges uniformly in x^- as $n \to \infty$, $0 \le i < N$.*

Proof. It will suffice to investigate the ratios on the right side of (3.8) for $x \in \Omega_A$. To avoid negative subscripts, define now

$$\mu_n(x) = \frac{\mu[x_0, \dots, x_n, x_{n+1}]}{\mu[x_0, \dots, x_n]}.$$

It has to be shown that

$$\lim_{s \to \infty} \sup\{|\mu_{s+k}(x) - \mu_{s+k}(y)| : x \in \Omega_A,\ y \in \Omega_A,\ x_i = y_i \text{ for } k \le i \le k+s+1,\ k \ge 0\} = 0. \tag{3.9}$$

It follows from the definition of a Gibbs measure that $\mu_n(x)$ is bounded away from 0 for $x \in \Omega_A$. For (3.9) more information about μ is needed. As shown in [1] there is a useful formula for $\mu[x_0, \dots, x_n]$ involving an auxiliary measure $\mu' \in \mathscr{M}^A_\theta$, a strictly positive function $h \in H_A$ such that $h(x) = h(y)$ if $x, y \in \Omega_A$, $x_i = y_i$ for $i \ge 0$. Since h does not depend on the past write $h(x_0 x_1, \dots)$ for $h(x)$. If $z = (z_0, z_1, \dots)$, then $(x_0, \dots, x_n z)$ denotes the sequence $(x_0, \dots, x_n, z_0, z_1, \dots)$. With $\lambda = e^P$ the formula for $\mu[x_0, \dots, x_n]$ reads

$$\mu[x_0, \dots, x_n] = \lambda^{-(n+1)} \int \exp\left\{ \sum_{j=0}^{m} \phi(\theta^j(x_0, \dots, x_n z)\right\} h(x_0, \dots, x_n z) \mu'(dz), \tag{3.10}$$

the integral extending over all $z = (z_0, z_1, \dots)$ such that $(x_0, \dots, x_n z)$ is A-admissible. Assume $(x_0, \dots, x_n)$ is admissible and let $w = (w_0, w_1, \dots)$ be any sequence such that $(x_0, \dots, x_n w)$ is admissible. Let $n = k + s$, $s \ge 1$. On the right side of (3.10) replace the argument $(x_0, \dots, x_n z)$ by $(x_0, \dots, x_{n-1} w)$ in those terms of the exponential sum corresponding to $0 \le j \le k$, and make the same replacement in the argument of h. One obtains

$$\lambda^{-(n+1)} \exp\left\{ \sum_{j=0}^{k} \phi(\theta^j(x_0, \dots, x_{n-1} w)\right\} h(x_0, \dots, x_{n-1} w)$$

$$\times \int \exp\left\{ \sum_{j=k+1}^{k+s} \phi(\theta^j(x_0, \dots, x_n z)\right\} \mu'(dz)$$

and the ratio of $\mu[x_0,\dots,x_n]$ to this quantity differs from 1 by a quantity going to 0 exponentially as $s\to\infty$. In like manner $\mu[x_0,\dots,x_{n-1}]$ is approximated by

$$\lambda^{-n}\exp\left\{\sum_{j=0}^{k}\phi(\theta^j(x_0,\dots,x_{n-1})\right\}h(x_0,\dots,x_{n-1}w)$$
$$\times\int\exp\left\{\sum_{j=k+1}^{k+s-1}\phi(\theta^j(x_0,\dots,x_nz)\right\}\mu'(dz).$$

Hence, defining for $n=k+s$,

$$\alpha_{n,s}(x)=\lambda^{-1}\int\exp\left\{\sum_{j=k+1}^{k+s}\phi(\theta^j(x_0,\dots,x_nz)\right\}\mu'(dz)$$
$$\times\left[\int\exp\left\{\sum_{j=k+1}^{k+s+1}\phi(\theta^j(x_0,\dots,x_{n-1}z)\right\}\mu'(dz)\right]^{-1},$$

one finds

$$\lim_{s\to\infty}\sup\left\{\left|\frac{\mu[x_0,\dots,x_n]}{\mu[x_0,\dots,x_{n-1}]}-\alpha_{n,s}(x)\right|:n>s,x\in\Omega_A\right\}=0.$$

Note that if $y_i=x_i$ for $k\le i\le k-s$, then $\alpha_{n,s}(x)=\alpha_{n,s}(y)$. So (3.9) is proved, and the lemma follows. □

PROOF OF THEOREM 3.2. Let μ be a Gibbs measure. Using Lemma 3.3, define

$$\mu^*_{\omega^-}(B_i)=\lim_{n\to\infty}\mu^{(n)}_{\omega^-}(i),$$

for $\omega\in\Omega_A$. Then the conditions of Proposition 3.1 hold and so the conclusion of this proposition applies. Now Theorems 1.1 and 2.1 apply. This justifies all parts of Theorem 3.2 except the second equality in (3.7).

For $\nu\in\mathscr{M}_\theta^A$ define

$$R_n^\nu(\omega)=\frac{1}{n}\log\frac{\nu[\omega_0,\dots,\omega_{n-1}]}{\mu[\omega_0,\dots,\omega_{n-1}]}.$$

By (3.6)

$$R_n^\nu(\omega)=\frac{1}{n}\log\nu[\omega_0,\dots,\omega_{n-1}]-\frac{1}{n}\sum_{k=0}^{n-1}\phi(\theta^k\omega)+P+\alpha_n(\omega),$$

where $\alpha_n(\omega)\to0$ uniformly in ω. One obtains immediately that as $n\to\infty$,

$$(\nu,R_n^\nu)\to-h_\nu-(\nu,\phi)+P. \tag{3.11}$$

We wish to identify the right side of (3.11) as $H(\nu;\mu^*)$. For $\omega\in\Omega_A$ put

$$f_{k,n}(\omega)=\frac{\nu[\omega_k,\dots,\omega_n]}{\mu[\omega_k,\dots,\omega_n]},\qquad r_n(\omega)=\frac{f_{-n,1}(\omega)}{f_{-n,0}(\omega)}$$

and note that if the denominator in r_n vanishes so does the numerator: In that case interpret the ratio arbitrarily (e.g., set it equal to 1). Now $R_n^\nu = n^{-1}\log f_{0,n-1}$. Note

$$\frac{1}{n}\log f_{0,n} = \frac{1}{n}\log f_{0,0} + \frac{1}{n}\sum_{j=0}^{n-1}\log r_j \circ \theta^j. \tag{3.12}$$

It will be shown that

$$(\nu, \log r_n) \to H(\nu; \mu^*) \tag{3.13}$$

and then (3.12) will imply $(\nu, n^{-1}\log f_{0,n}) \to H(\nu; \mu^*)$, as desired.

Recall the notation introduced in (3.8); $\nu_{x^-}^{(n)}(i)$ will have the corresponding definition, where here and in the subsequent discussion 0/0 may be interpreted to have the value 1. Also (ν_{ω^-}) will denote regular conditional probabilities $\nu[\ |\mathscr{F}_{-\infty,0}](\omega)$. Write

$$r_n(\omega) = \frac{\nu[\omega_{-n},\dots,\omega_0,\omega_1]}{\mu[\omega_{-n},\dots,\omega_0,\omega_1]}\frac{\mu[\omega_{-n},\dots,\omega_0]}{\nu[\omega_{-n},\dots,\omega_0]} = \frac{\nu_{\omega^-}^{(n)}(\omega_1)}{\mu_{\omega^-}^{(n)}(\omega_1)}$$

and so

$$\begin{aligned}(\nu, \log r_n) &= \int \nu_{\omega^-}(\log r_n)\nu(d\omega^-)\\ &= \int\left(\sum_{i=0}^{N-1}\nu_{\omega^-}^{(n)}(i)\log \nu_{\omega^-}^{(n)}(i)\right)\nu(d\omega^-)\\ &\quad - \int\left(\sum_{i=0}^{N-1}\nu_{\omega^-}^{(n)}(i)\log \mu_{\omega^-}^{(n)}(i)\right)\nu(d\omega^-).\end{aligned} \tag{3.14}$$

By the martingale convergence theorem $\nu_{\omega^-}^{(n)}(i) \to (\nu_{\omega^-})_1(i)$, ν a.e., and by Lemma 3.3 $\mu_{\omega^-}^{(n)}(i) \to (\mu_{\omega^-}^*)_1(i)$ uniformly in ω^-. It follows then from (3.14) that $(\nu, \log r_n) \to H(\nu; \mu^*)$. This justifies (3.7) for $\nu \in \mathscr{M}_\theta^A$. If $\nu \in \mathscr{M}_\theta \setminus \mathscr{M}_\theta^A$, then for some k, $\nu_{0,k}$ is not absolutely continuous with respect to $\mu_{0,k}$. By the basic properties of entropy (see [2]) this implies $H(\nu; \mu^*) = \infty$ and again (3.7) holds. □

For the class of Gibbs measures treated here one can actually verify a stronger condition than (RM). Namely, there exists a constant c such that

$$\log\frac{\mu_{\eta^-}^*(A)}{\mu_{\omega^-}^*(A)} < c, \quad \text{for } \eta \in \Omega^-,\ \omega \in \Omega^-,\ A \in \mathscr{F}_{m,n},\ n \geq m \geq 0.$$

4. Examples and remarks. The following examples and remarks give some more insight into the applicability of large deviation principles to stationary shifts.

EXAMPLE 4.1 (Adapted from Sokal [6]). This example will present a strongly mixing shift on $M = \{-1, +1\}$ for which the large deviation principle fails. According to a basic result of Donsker and Varadhan (cf. [2], Corollary 1.7) if

$\mu \in \mathscr{M}_\theta$ is a stationary shift obeying a large deviation principle with rate function K, then for any bounded and continuous random variable Y,

$$\lim_{n\to\infty} \frac{1}{n}\log \mu\left(\exp\left\{\sum_{k=0}^{n-1} Y\circ\theta^k\right\}\right) = \sup_{\nu\in\mathscr{M}} \{\langle \nu, Y\rangle - K(\nu)\}. \tag{4.1}$$

In the present example the limit in (4.1) will fail to exist for $Y(\omega) = \omega_0$. It remains to specify $\mu \in \mathscr{M}_\theta$. The measure μ is concentrated on those ω consisting of a sequence of *blocks* of $+1$'s followed by an equal number of -1's. The *length* of each block is the same as the number of $+1$'s and -1's, (i.e., twice the number of $+1$'s). It is assumed that the lengths of successive blocks form a sequence of independent identically distributed random variables. Thus any time coordinate, say 0, belongs to a block of random length $2L$. Let $a_n = \mu[L = n]$. Of course, since μ is to be stationary, given that $L = n$, the coordinate 0 is equally likely to occupy the first, second, ..., $2n$th position of the block. Let N be an integer, $N > 2$, and choose $0 < \alpha < (N-1)/N$. Let $n_i = N^i$, $i = 1, 2 \dots$, and let $a(n_i) = c_\alpha e^{-\alpha n_i}$, $a(n) = 0$, if $n \neq N^i$ for every i, with c_α so that $\Sigma a(n) = 1$. This determines $\mu \in \mathscr{M}_\theta$. Now let $S_k = \Sigma_{i=0}^{k-1}\omega_i$, $g(k) = \mu(\exp\{S_k\})$. Our claim is that as $n \to \infty$, $n^{-1}\log g(n)$ oscillates. To see this, note $\mu[S_{n_i} = n_i] = P[0$ occupies the first position of a block of length $2n_i] = a(n_i)/(2n_i)$, and this implies

$$g(n_i) > c_\alpha(2n_i)^{-1}\exp\{n_i(1-\alpha)\}. \tag{4.2}$$

Next consider $k_i = \beta n_i$, where $1 < \beta < N$. If, for given ω, k_i is in a block of length $2n$, with $n < n_i$, then $S_{k_i} < n < n_i$; if k_i is in a block of length $2n$ with $n > n_{i+1}$ use $S_{k_i} < k_i$. Note there exists c'_α such that $\mu[k_i$ in a block of length $> 2n_{i+1}] < c'_\alpha\exp\{-\alpha n_{i+1}\}$. Hence

$$\mu\left(\exp\{S_{k_i}\}\right) < \exp\{n_i\} + c'_\alpha\exp\{-\alpha n_{i+1} + \beta n_i\}.$$

Now choose $\beta = N\alpha + 1$ and find

$$g(k_i) < (1 + c'_\alpha)e^{n_i} = (1 + c'_\alpha)\exp\{k_i/\beta\}. \tag{4.3}$$

Our assumptions ensure $\beta^{-1} < (1-\alpha)$ and oscillation is established.

EXAMPLE 4.2. Consider a shift $(\Omega(M), \mathscr{F}, \theta, \mu)$, with $\mu \in \mathscr{M}_\theta$ such that the (ω_i) are m-dependent [i.e., $(\dots\omega_{n-1}, \omega_n)$ and $(\omega_{n+m+1}, \omega_{n+m+2}, \dots)$ are independent]. Then (RM) certainly holds. However, easy examples show (even for $M = \{0, 2, 2\}$) that (CD) may fail.

EXAMPLE 4.3. Piecewise monotonic maps of the unit interval onto itself have been extensively studied. In [4] it was shown that under familiar hypotheses these maps have an invariant measure and the corresponding symbolic dynamics give rise to a stationary shift satisfying (RM). Thus Theorem 1.1 gives the existence of a rate function. From [8] an interesting expression for the rate function becomes available.

EXAMPLE 4.4. Let μ be the Bernoulli shift on $\Omega(M)$, $M = \{0, 1\}$, with $\mu[\omega_0 = 0] = \mu[\omega_0 = 1] = \frac{1}{2}$. Define (η_i) by $\eta_i = \sum_{k=1}^{\infty} 2^{-k}\omega_{i+k}$. This gives rise to a new shift on [0, 1], for which the large deviation principle will hold; this follows from the Donsker–Varadhan contraction principle, or see the discussion in [4]. This shift is deterministic (η_0 determines all η_i, with $i > 0$), and (RM) does not hold. This indicates that it may be difficult to find necessary conditions for the large deviation principle.

REMARK 4.5. Instead of the condition (CD), which gives a continuous dependence on the past, one can condition on the future and ask for continuous dependence. This is the approach of Takahashi [8]. In the case of a Gibbs measure μ for $\phi \in H_A$ it follows from the formula (3.10) that

$$\begin{aligned} \frac{\mu[x_0, x_1, \ldots, x_n]}{\mu[x_1, \ldots, x_n]} &\to j_\mu(x_0 x_1, \ldots) \\ &= \frac{h(x_0 x_1, \ldots)}{h(x_1, \ldots)} \lambda^{-1} \exp\{\phi(x_0 x_1, \ldots)\} \end{aligned} \tag{4.4}$$

as $n \to \infty$ uniformly for $x \in \Omega_A$. So j_μ is continuous and it follows from [9] that

$$K(\nu) = -h_\nu + (\nu, -\log j_\mu).$$

Indeed one can extend Theorem 3.2 to a wider class of Gibbs measures if one uses the results of [5], Exercise 2, page 97.

REFERENCES

[1] BOWEN, R. (1975). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Math.* **470**. Springer, Berlin.

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

[3] DONSKER, M. and VARADHAN, S. R. S. (1985). Large deviations for stationary Gaussian processes. *Comm. Math. Phys.* **97** 187–210.

[4] OREY, S. (1984). Large deviations in ergodic theory. In *Seminar on Stochastic Processes, 1984* (E. Çinlar, K. L. Chung and R. K Getoor, eds.) 195–249. Birkhäuser, Boston.

[5] RUELLE, D. (1978). *Thermodynamic Formalism.* Addison-Wesley, Reading, Mass.

[6] SOKAL, A. (1982). Private communication.

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

[8] TAKAHASHI, Y. (1982). Entropy functional (free energy) for dynamical systems and their random perturbations. In *Proc. Taniguchi Symp. on Stochastic Anal. at Katata and Kyoto, 1982.* Kinokuniga, Tokyo/North-Holland, Amsterdam.

[9] TAKAHASHI, Y. (1984). Variational principle of Gibbs type and large deviation theory. Preprint.

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

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CINCINNATI
CINCINNATI, OHIO 45267