LARGE DEVIATIONS FOR THE EMPIRICAL MEASURE OF A MARKOV CHAIN WITH AN APPLICATION TO THE MULTIVARIATE EMPIRICAL MEASURE[1]

By Richard S. Ellis

University of Massachusetts

The main theorems in this paper prove uniform large deviation properties for the empirical measure and the multivariate empirical measure of a Markov chain that takes values in a complete separable metric space. One contribution of the paper is that, in contrast to previous large deviation results for the empirical measure, we do not assume that the transition probability of the Markov chain has a density with respect to a reference measure.

1. Introduction. Let $X_0, X_1, X_2, \ldots$ be a Markov chain on a space Ω with stationary transition probabilities $\pi(x, dy)$. We assume that the Markov chain takes values in a complete separable metric space $\mathscr{X}$. For each starting point $X_0 = x \in \mathscr{X}$, π induces a probability measure P_x on Ω. For each $\omega \in \Omega$, positive integer n and Borel set A in $\mathscr{X}$, define the empirical measure

$$L_n(\omega, A) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{X_j(\omega)}\{A\}.$$

We denote by $\mathscr{M}(\mathscr{X})$ the space of Borel probability measures on $\mathscr{X}$ with the topology of weak convergence. For each positive integer n, L_n maps Ω into $\mathscr{M}(\mathscr{X})$.

The pioneering work on large deviations for the empirical measure of a Markov chain is that of Donsker and Varadhan (1975, 1976). Our first theorem, Theorem 1.2, and the work of Stroock (1984), on which it is based, prove uniform versions of results in Donsker and Varadhan (1976). As an application of Theorem 1.2, we derive in Theorem 1.4 a uniform large deviation property for a natural generalization of the empirical measure, called the multivariate empirical measure. Theorem 1.4 cannot be proved from the special case of Theorem 1.2 given in Stroock (1984).

We next state the hypotheses needed for the large deviation theorems.

HYPOTHESIS 1.1(a). $\pi(x, dy)$ is a Feller transition probability on $\mathscr{X}$.

HYPOTHESIS 1.1(b).   For some $\beta \in \mathbb{Z}^+$, some $M \in [1, \infty)$, all $x, x' \in \mathscr{X}$ and all Borel sets A in $\mathscr{X}$,

$$\pi^\beta(x, A) \le M\pi^\beta(x', A),$$

where $\pi^\beta(x, dy)$ denotes the β-step transition probability of the Markov chain.

The next theorem, Theorem 1.2, is proved in Stroock (1984) in the case where $\pi(x, dy)$ satisfies Hypotheses 1.1(a)–(b) for $\beta = 1$. The extension to $\beta \geq 2$ is new. Theorem 1.2 is proved in Section 2.

The large deviation theorems in Donsker and Varadhan (1975, 1976) require that $\pi(x, dy)$ have a density $\pi(x, y)$ with respect to a reference measure $\lambda(dy)$. The large deviation theorems in the present paper do not require the existence of a density of $\pi(x, dy)$.

THEOREM 1.2.   *Let $\mathscr{X}$ be a complete separable metric space. We assume that* $\pi(x, dy)$ *satisfies Hypotheses 1.1(a)–(b). Define for* $\mu \in \mathscr{M}(\mathscr{X})$,

$$J_\pi(\mu) = \sup_{u \in \mathscr{U}(\mathscr{X})} \int_{\mathscr{X}} \frac{u(x)}{(\pi u)(x)} \mu(dx),$$

where $\mathscr{U}(\mathscr{X})$ denotes the set of $u \in \mathscr{C}(\mathscr{X})$ such that $u \geq \varepsilon$ on $\mathscr{X}$ for some $\varepsilon = \varepsilon(u) > 0$. The following conclusions hold.

(a) $J_\pi(\mu)$ *is a lower semicontinuous convex function of* $\mu \in \mathscr{M}(\mathscr{X})$.
(b) *For any $L \geq 0$, the set $\{\mu \in \mathscr{M}(\mathscr{X}) : J_\pi(\mu) \leq L\}$ is compact.*
(c) *For each closed set F in $\mathscr{M}(\mathscr{X})$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathscr{X}} P_x\{\omega : L_n(\omega, \cdot) \in F\} \right) \leq -\inf_{\mu \in F} J_\pi(\mu).$$

(d) *For each open set G in $\mathscr{M}(\mathscr{X})$,*

$$\liminf_{n \to \infty} \frac{1}{n} \log\left(\inf_{x \in \mathscr{X}} P_x\{\omega : L_n(\omega, \cdot) \in G\} \right) \geq -\inf_{\mu \in G} J_\pi(\mu).$$

REMARK 1.3.   *Extensions.* With additional work one may show that the conclusions of Theorem 1.2 remain valid under the following weaker hypotheses.

(a) Hypothesis 1.1(a) may be dropped; i.e., $\pi(x, dy)$ need not be a Feller transition probability.
(b) Hypothesis 1.1(b) may be replaced by the following weaker hypothesis.

HYPOTHESIS 1.1(b′).   For some β and N in $\mathbb{Z}^+$, some $M \in [1, \infty)$, all $x, x' \in \mathscr{X}$ and all Borel sets A in $\mathscr{X}$,

$$(1.1) \qquad\qquad \pi^\beta(x, A) \leq \frac{M}{N} \sum_{i=1}^{N} \pi^i(x', A).$$

In particular, Theorem 1.2 is valid for a finite-state Markov chain that is irreducible.

Remark 2.1 at the end of Section 2 will indicate why these extensions of Theorem 1.2 hold.

We now apply Theorem 1.2 to prove a uniform large deviation property for the *multivariate empirical measure*. The latter is defined for each $\omega \in \Omega$,

positive integer n, positive integer $\alpha \geq 2$ and Borel set A in the product space $\mathscr{X}^\alpha$ by

$$M_{n,\alpha}(\omega, A) = \frac{1}{n} \sum_{j=0}^{n-1} \delta_{Y_{j,\alpha}(\omega)}\{A\},$$

where

$$Y_{j,\alpha}(\omega) = \left(X_j(\omega), X_{j+1}(\omega), \ldots, X_{j+\alpha-1}(\omega)\right) \in \mathscr{X}^\alpha,$$

i.e., $Y_{j,\alpha}(\omega$ is that element of $\mathscr{X}^\alpha$ with coordinates

$$\left(Y_{j,\alpha}(\omega)\right)_k = X_{j+k-1}, \quad \text{for } 1 \leq k \leq \alpha.$$

For each positive integer n and each positive integer $\alpha \geq 2$, $M_{n,\alpha}$ maps Ω into $\mathscr{M}(\mathscr{X}^\alpha)$, the space of Borel probability measures on $\mathscr{X}^\alpha$.

The multivariate empirical measure arises naturally in a number of contexts in statistical mechanics and in statistics. For the former, see Ellis (1985), especially Appendix C.6, and Szulga, Woyczynski, Mann and Tjatjopoulos (1987); for the latter, see Example 5 in Section 2.2 of Pollard (1984).

In order to state Theorem 1.4, we need some definitions. Denote by $\hat{X}_j$, $j \in \{1, 2, \ldots, \alpha\}$, the coordinate functions on $\mathscr{X}^\alpha$; i.e., for $\mathbf{x} = (x_1, x_2, \ldots, x_\alpha) \in \mathscr{X}^\alpha$, $\hat{X}_j(\mathbf{x}) = x_j$. Let μ be a Borel probability measure on $\mathscr{X}^\alpha$. We denote by $\mu^*(x_1, \ldots, x_{\alpha-1}, dx_\alpha)$ a regular conditional distribution of $\hat{X}_\alpha$ given $\hat{X}_1 = x_1, \ldots, \hat{X}_{\alpha-1} = x_{\alpha-1}$. We denote by $\mu_1(dx_1 \times \cdots \times dx_{\alpha-1})$ the $(\alpha - 1)$-dimensional marginal of μ which is the distribution of $\hat{X}_1, \ldots, \hat{X}_{\alpha-1}$ and by $\mu_\alpha(dx_2 \times \cdots \times dx_\alpha)$ the $(\alpha - 1)$-dimensional marginal of μ which is the distribution of $\hat{X}_2, \ldots, \hat{X}_\alpha$. Given ν and ρ Borel probability measures on $\mathscr{X}$, we define the relative entropy of ν with respect to ρ,

$$I(\nu, \rho) = \int_{\mathscr{X}} \log \frac{d\nu}{d\rho}(x)\nu(dx), \quad \text{if } \nu \ll \rho,$$

$$= +\infty, \qquad\qquad\qquad \text{otherwise.}$$

The following theorem is proved in Section 3 of this paper by applying Theorem 1.2 to the process $\{Y_{j,\alpha}, j = 0, 1, \ldots\}$, which is a Markov chain with state space $\mathscr{X}^\alpha$.

THEOREM 1.4. *Let $\mathscr{X}$ be a complete separable metric space. We assume that $\pi(x, dy)$ satisfies Hypotheses 1.1(a)–(b). For $\alpha \in \{2, 3, \ldots\}$ and μ a Borel probability measure on $\mathscr{X}^\alpha$, define*

$$I_{\pi,\alpha}(\mu) = \int_{\mathscr{X}^{\alpha-1}} I(\mu^*(x_1, \ldots, x_{\alpha-1}, \cdot), \pi(x_{\alpha-1}, \cdot))\mu_1(dx_1 \times \cdots \times dx_{\alpha-1}),$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if } \mu_1 = \mu_\alpha,$$

$$= +\infty, \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textit{if } \mu_1 \neq \mu_\alpha.$$

Denote by $\mu_1 \otimes \pi$ the Borel probability measure on $\mathscr{X}^\alpha$ defined by

$$\mu_1 \otimes \pi(dx_1 \times \cdots \times dx_\alpha) = \mu_1(dx_1 \times \cdots \times dx_{\alpha-1})\pi(x_{\alpha-1}, dx_\alpha).$$

The following conclusions hold.

(a) *$I_{\pi,\alpha}(\mu)$ is a lower semicontinuous convex function of $\mu \in \mathscr{M}(\mathscr{X}^\alpha)$.*

(b) *For any $L \geq 0$, the set $\{\mu \in \mathscr{M}(\mathscr{X}^\alpha): I_{\pi,\alpha}(\mu) \leq L\}$ is compact.*

(c) *For each closed set F in $\mathscr{M}(\mathscr{X}^\alpha)$,*

$$\limsup_{n \to \infty} \frac{1}{n}\log\left(\sup_{x \in \mathscr{X}} P_x\{\omega: M_{n,\alpha}(\omega, \cdot) \in F\}\right) \leq -\inf_{\mu \in F} I_{\pi,\alpha}(\mu).$$

(d) *For each open set G in $\mathscr{M}(\mathscr{X}^\alpha)$,*

$$\liminf_{n \to \infty} \frac{1}{n}\log\left(\inf_{x \in \mathscr{X}} P_x\{\omega: M_{n,\alpha}(\omega, \cdot) \in G\}\right) \geq -\inf_{\mu \in G} I_{\pi,\alpha}(\mu).$$

(e) *For all $\mu \in \mathscr{M}(\mathscr{X}^\alpha)$,*

$$I_{\pi,\alpha}(\mu) = I(\mu, \mu_1 \otimes \pi).$$

REMARK 1.5. *Extensions.* As in the case of Theorem 1.2, the conclusions of Theorem 1.4 remain valid under the weaker hypotheses given in Remark 1.3(a)–(b). Remark 3.5 at the end of Section 3 will indicate why this is true.

In the case of finite-state Markov chains, large deviations for the multivariate empirical measure have been studied by a number of authors, including Csiszár, Cover and Choi (1987), Ellis (1985) and Natarajan (1985, 1986). In future work, we will apply Theorem 1.4 to prove a uniform large deviation property for the empirical process of a Markov chain that takes values in a complete separable metric space [Ellis and Wyner (1987)]. The proof is based on an approximation argument that uses the uniform large deviation property of the multivariate empirical measure.

2. Proof of Theorem 1.2. Theorem 1.2 is proved in Stroock (1984) under the assumption that $\pi(x, dy)$ satisfies Hypotheses 1.1(a)–(b) with $\beta = 1$. The proof consists of two parts. Part 1, given in Section 6, shows the existence of a rate function $I_\pi(\mu)$ such that the P_x-distributions of $L_n(\omega, \cdot)$ have a uniform large deviation property with this rate function. Part 2, given in Section 7, identifies $I_\pi(\mu)$ with $J_\pi(\mu)$. We now indicate the minor modifications needed in order to prove Part 1 under the assumption that $\pi(x, dy)$ satisfies Hypotheses 1.1(a)–(b) with $\beta \geq 2$. Part 2, which proves that $I_\pi(\mu)$ equals $J_\pi(\mu)$, then follows as in Theorem 7.18 and Corollary 7.21 of Stroock (1984).

Given positive integers n and i and a point $\omega \in \Omega$, we define

$$L_n^i(\omega, \cdot) = \frac{1}{n}\sum_{j=i}^{n-1+i} \delta_{X_j(\omega)}\{\cdot\}$$

and denote by $Q_{n,x}^i$ the P_x-distribution of $L_n^i(\omega, \cdot)$. Hypothesis 1.1(b) implies

that

(2.1)
$$\sup_{x \in \mathcal{X}} Q_{n,x}^{\beta}\{\cdot\} \le M \inf_{x \in \mathcal{X}} Q_{n,x}^{\beta}\{\cdot\}.$$

Define $\mathscr{P}_n\{\cdot\} = \inf_{x \in \mathcal{X}} Q_{n,x}^1\{\cdot\}$. As in the proof of Lemma 6.4 in Stroock (1984), for any convex set A in $\mathscr{M}(\mathcal{X})$ and all positive integers m and n,

(2.2)
$$\mathscr{P}_{m+n}\{A\} \ge \mathscr{P}_m\{A\} \cdot \mathscr{P}_n\{A\}.$$

However, we must modify the other arguments given on page 117 of Stroock (1984) since it is no longer clear that if A is open and convex, then $\mathscr{P}_m\{A\} > 0$ for some $m \in \mathbb{Z}^+$ implies that $\mathscr{P}_n\{A\} > 0$ for all sufficiently large $n \in \mathbb{Z}^+$. Analogous complications arise in Stroock's treatment of continuous time processes (see pages 121–131).

Let Φ denote the Lévy–Prohorov metric on $\mathscr{M}(\mathcal{X})$. It follows from (2.2) that if A is convex and $\mathscr{P}_m\{A\} > 0$ for some $m \in \mathbb{Z}^+$, then there exists $\delta > 0$ such that $\mathscr{P}_n\{A^{(\delta)}\} > 0$ for all sufficiently large $n \in \mathbb{Z}^+$, where

$$A^{(\delta)} = \{\mu \in \mathscr{M}(\mathcal{X}) \colon \exists\, \nu \in A \text{ satisfying } \Phi(\mu, \nu) < \delta\}.$$

For $\mu \in \mathscr{M}(\mathcal{X})$ and $r > 0$, we denote by $B(\mu, r)$ the open ball

$$B(\mu, r) = \{\nu \in \mathscr{M}(\mathcal{X}) \colon \Phi(\mu, \nu) < r\}.$$

It follows that for $\mu \in \mathscr{M}(\mathcal{X})$ and $r > 0$, the quantity

(2.3)
$$\mathscr{L}(\mu, r) = -\lim_{n \to \infty} \frac{1}{n}\log \mathscr{P}_n\{B(\mu, r)\}, \quad \text{if } \sup_{n \ge 1}\mathscr{P}_n\{B(\mu, r/2)\} > 0,$$

$$= \infty, \qquad\qquad\qquad\qquad\qquad \text{otherwise,}$$

is well defined. For $\mu \in \mathscr{M}(\mathcal{X})$, we define

(2.4)
$$I_\pi\{\mu\} = \sup_{r > 0}\mathscr{L}(\mu, r) = \lim_{r \downarrow 0}\mathscr{L}(\mu, r).$$

Part 1 of the proof of Theorem 1.2 consists in showing that the conclusions of Theorem 1.2 hold with $J_\pi(\mu)$ replaced by $I_\pi(\mu)$.

(a) That $I_\pi(\mu)$ is a lower semicontinuous convex function of $\mu \in \mathscr{M}(\mathcal{X})$ is proved as in Lemmas 3.15 and 6.12 of Stroock (1984).

(d) The uniform lower large deviation bound with rate function $I_\pi(\mu)$ is proved as on pages 118–119 and 128–129 of Stroock (1984).

(b) First assume that $\pi(x, dy)$ satisfies Hypothesis 1.1(b) for $\beta = 2$. For $i = 0, 1$ we define the processes

$$L_{n,i}^2(\omega, \cdot) = \frac{1}{\theta_i(n)}\sum_{j=1}^{\theta_i(n)}\delta_{X_{2j+1}(\omega)}\{\cdot\},$$

where if n is odd, then $\theta_0(n) = (n+1)/2$ and $\theta_1(n) = (n-1)/2$, whereas if n is even, then $\theta_0(n) = \theta_1(n) = n/2$. Just as in the proof of Lemma 6.5 in Stroock (1984), one can show that for each $L > 0$ there exists a compact set K_L in

$\mathcal{M}(\mathcal{X})$ such that K_L is convex and for each $i = 0, 1$,

$$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} P_x\{L_{n,1}^2 \in K_L^c\} \right) \leq -L.$$

These bounds imply that

$$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} Q_{n,x}^2\{K_L^c\} \right) \leq -L.$$

In a similar way one shows that if $\pi(x, dy)$ satisfies Hypothesis 1.1(b) for $\beta \geq 3$, then for each $L > 0$ there exists a compact set K_L in $\mathcal{M}(\mathcal{X})$ such that K_L is convex and for each $i = 0, 1$,

(2.5) $$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} Q_{n,x}^\beta\{K_L^c\} \right) \leq -L.$$

Since $\inf_{x \in \mathcal{X}} Q_{n,x}^\beta\{\cdot\} \geq \inf_{x \in \mathcal{X}} Q_{n,x}\{\cdot\}$, the proof that I_π has compact level sets follows the proof on page 119 of Stroock (1984).

(c) Using (2.1), we may prove as on pages 120 and 129–130 of Stroock (1984) that for F a nonempty compact set in $\mathcal{M}(\mathcal{X})$,

(2.6) $$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} Q_{n,x}^\beta\{F\} \right) \leq -\inf_{\mu \in F} I_\pi(\mu).$$

Since for each $L > 0$ there exists a compact set K_L in $\mathcal{M}(\mathcal{X})$ such that (2.5) holds, (2.6) is also valid for any nonempty closed set F in $\mathcal{M}(\mathcal{X})$. For $\delta > 0$, let $\overline{F^{(\delta)}}$ denote the closure of the set $F^{(\delta)}$. Since

$$\sup_{\omega \in \Omega} \Phi\left(L_n^1(\omega, \cdot), L_n^\beta(\omega, \cdot) \right) < \frac{2(\beta - 1)}{n},$$

we have

$$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} Q_{n,x}\{F\} \right) \leq \limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} Q_{n,x}^\beta\{\overline{F^{(\delta)}}\} \right)$$

and so it follows from (2.6) that

$$\limsup_{n \to \infty} \frac{1}{n} \log\left(\sup_{x \in \mathcal{X}} Q_{n,x}\{F\} \right) \leq \lim_{\delta \downarrow 0}\left[-\inf\{I_\pi(\mu) : \mu \in \overline{F^{(\delta)}}\} \right] = -\inf_{\mu \in F} I_\pi(\mu).$$

This completes the proof of Theorem 1.2. □

REMARK 2.1. The purpose of this remark is to point out why the conclusions of Theorem 1.2 remain valid under the weaker hypotheses given in Remark 1.3(a)–(b). As we have seen, Theorem 1.2 is proved in two parts. Part 1 shows the existence of a rate function $I_\pi(\mu)$ such that the P_x-distributions of $L_n(\omega, \cdot)$ have a uniform large deviation property with this rate function. Part 2 identifies $I_\pi(\mu)$ with $J_\pi(\mu)$. We used Hypothesis 1.1(b) in the proof of Part 1. With some obvious modifications in this proof, one may easily show Part 1 under the weaker Hypothesis 1.1(b'). For the proof of Part 2, we referred to Stroock (1984). There,

the Feller property of $\pi(x, dy)$ together with Part 1 is used in order to identify the form of the rate function (see his Theorem 7.18 and Corollary 7.21). According to an observation of Stroock (1987), if $\pi(x, dy)$ does not have the Feller property, then Theorem 7.18 remains true by a modified proof that is based on an approximation argument. Together with Part 1, Theorem 7.18 yields Corollary 7.21, and thus Part 2 follows.

3. Proof of Theorem 1.4. The multivariate empirical measure $M_{n,\,\alpha}(\omega, \cdot)$ equals the empirical measure for the process

$$Y_{j,\,\alpha}(\omega) = \left(X_j(\omega), X_{j+1}(\omega), \ldots, X_{j+\alpha-1}(\omega) \right), \qquad j = 0, 1, \ldots \,.$$

The process $\{Y_{j,\,\alpha}\}$ is a Markov chain that takes values in the complete separable metric space $\mathscr{X}^\alpha$ and has stationary transition probabilities $\pi_\alpha(\mathbf{x}, d\mathbf{y})$ given by

$$\pi_\alpha(x_1, \ldots, x_\alpha, dy_1 \times \cdots \times dy_\alpha)$$

$$= P\{X_{j+1} \in dy_1, \ldots, X_{j+\alpha} \in dy_\alpha | X_j = x_1, \ldots, X_{j+\alpha-1} = x_\alpha\}$$

$$= \delta_{x_2}(dy_1)\delta_{x_3}(dy_2) \cdots \delta_{x_\alpha}(dy_{\alpha-1})\pi(x_\alpha, dy_\alpha).$$

For each starting point $Y_{0,\,\alpha} = \mathbf{x} \in \mathscr{X}^\alpha$, π_α induces a probability measure $\overline{P}_{\mathbf{x}}$ on Ω. In addition, for each point $\mathbf{x} \in \mathscr{X}^\alpha$, the underlying transition probability π on $\mathscr{X}$ induces a probability measure $P_{\mathbf{x}}$ on Ω such that $P_{\mathbf{x}}\{(X_0, X_1, \ldots, X_{\alpha-1}) = \mathbf{x}\} = 1$. It is clear that for any positive integer n and bounded measurable $F: (\mathscr{X}^\alpha)^n \to [0, \infty)$,

$$(3.1) \quad E^{\overline{P}_{\mathbf{x}}}\left\{F\left(Y_{0,\,\alpha}, Y_{1,\,\alpha}, \ldots, Y_{n-1,\,\alpha}\right)\right\} = E^{P_{\mathbf{x}}}\left\{F\left(Y_{0,\,\alpha}, Y_{1,\,\alpha}, \ldots, Y_{n-1,\,\alpha}\right)\right\}.$$

In order to prove Theorem 1.4, we will first check that π_α satisfies Hypotheses 1.1(a)–(b). It will follow from Theorem 1.2 and (3.1) that the $P_{\mathbf{x}}$-distributions of $\{M_{n,\,\alpha}\}$ have a uniform large deviation property with the convex rate function

$$J_{\pi_\alpha}(\mu) = \sup_{u \in \mathscr{U}(\mathscr{X}^\alpha)} \int_{\mathscr{X}^\alpha} \log \frac{u(x)}{(\pi_\alpha u)(x)} \mu(dx), \qquad \mu \in \mathscr{M}(\mathscr{X}^\alpha).$$

Since

$$\inf_{\mathbf{x} \in \mathscr{X}^\alpha} E^{P_{\mathbf{x}}}\left\{F\left(Y_{0,\,\alpha}, \ldots, Y_{n-1,\,\alpha}\right)\right\} \leq \inf_{x \in \mathscr{X}} E^{P_x}\left\{F\left(Y_{0,\,\alpha}, \ldots, Y_{n-1,\,\alpha}\right)\right\}$$

$$\leq \sup_{x \in \mathscr{X}} E^{P_x}\left\{F\left(Y_{0,\,\alpha}, \ldots, Y_{n-1,\,\alpha}\right)\right\}$$

$$\leq \sup_{\mathbf{x} \in \mathscr{X}^\alpha} E^{P_{\mathbf{x}}}\left\{F\left(Y_{0,\,\alpha}, \ldots, Y_{n-1,\,\alpha}\right)\right\},$$

it will follow that the P_x-distributions of $M_{n,\,\alpha}(\omega, \cdot)$ have a uniform large deviation property with the convex rate function $J_{\pi_\alpha}(\mu)$. We will then identify $J_{\pi_\alpha}(\mu)$ with the function $I_{\pi,\,\alpha}(\mu)$ and with the relative entropy $I(\mu, \mu_1 \otimes \pi)$ defined in Theorem 1.4.

The proof of Theorem 1.4 involves four steps.

Step 1. π_α is a Feller transition probability on $\mathscr{X}^\alpha$ [Hypothesis 1.1(a)].

Step 2. For all $x, x' \in \mathscr{X}^\alpha$ and all Borel sets A in $\mathscr{X}^\alpha$,

$$\pi_\alpha^{\alpha+\beta-1}(\mathbf{x}, A) \le M\pi_\alpha^{\alpha+\beta-1}(\mathbf{x}', A)$$

[Hypothesis 1.1(b)].

Step 3. For all $\mu \in \mathscr{M}(\mathscr{X}^\alpha)$, $J_{\pi_\alpha}(\mu) = I_{\pi,\alpha}(\mu)$.

Step 4. For all $\mu \in \mathscr{M}(\mathscr{X}^\alpha)$, $I_{\pi,\alpha}(\mu) = I(\mu, \mu_1 \otimes \pi)$.

Step 1 is an elementary calculation and we omit the details. Step 2 follows from Hypothesis 1.1(b) and the fact that for any $n \in \mathbb{Z}^+ \cup \{0\}$,

$$\pi_\alpha^{\alpha+n}(x_1, \ldots, x_\alpha, dy_1 \times \cdots \times dy_\alpha) = \pi^{n+1}(x_\alpha, dy_1)\pi(y_1, dy_2) \cdots \pi(y_{\alpha-1}, dy_\alpha).$$

Step 4 is proved by routine manipulations involving Radon–Nikodym derivatives and we omit the details. The heart of the proof is Step 3, which we carry out for $\alpha = 2$. The proof for arbitrary $\alpha \ge 3$ follows by obvious changes in notation. We show that for each $\mu \in \mathscr{M}(\mathscr{X}^2)$,

$$J_{\pi_2}(\mu) = \sup_{u \in \mathscr{U}(\mathscr{X}^2)} \int_{\mathscr{X}^2} \log \frac{u(x)}{(\pi_2 u)(x)} \mu(dx)$$

equals

$$I_{\pi,2}(\mu) = \int_{\mathscr{X}} I(\mu^*(x_1, \cdot), \pi(x_1, \cdot))\mu_1(dx_1), \quad \text{if } \mu_1 = \mu_2,$$

$$= \infty, \qquad\qquad\qquad\qquad \text{if } \mu_1 \ne \mu_2.$$

Since $\pi_2(x_1, x_2, dy_1 \times dy_2) = \delta_{x_2}(dy_1)\pi(x_2, dy_2)$,

$$J_{\pi_2}(\mu) = \sup_{u \in \mathscr{U}(\mathscr{X}^2)} \left\{ \int_{\mathscr{X}^2} \log u(x_1, x_2)\mu(dx_1 \times dx_2) \right.$$

$$\left. - \int_{\mathscr{X}^2} \left[\log \int_{\mathscr{X}} u(x_2, y)\pi(x_2, dy) \right] \mu(dx_1 \times dx_2) \right\}$$

$$= \sup_{u \in \mathscr{U}(\mathscr{X}^2)} \left\{ \int_{\mathscr{X}} \left[\int_{\mathscr{X}} \log u(x_1, x_2)\mu^*(x_1, dx_2) \right] \mu_1(dx_1) \right.$$

$$\left. - \int_{\mathscr{X}} \left[\log \int_{\mathscr{X}} u(x_2, y)\pi(x_2, dy) \right] \mu_2(dx_2) \right\}.$$

We begin the proof of Step 3 with the next lemma.

LEMMA 3.1. *If $\mu_1 \ne \mu_2$, then $J_{\pi_2}(\mu) = +\infty$.*

PROOF. Let $u(x_1, x_2) = v(x_1)$, where v is any function in $\mathscr{U}(\mathscr{X})$. Then

$$(3.2) \qquad J_{\pi_2}(\mu) \ge \int_{\mathscr{X}} \log v(x_1)\mu_1(dx_1) - \int_{\mathscr{X}} \log v(x_2)\mu_2(dx_2).$$

If $\mu_1 \ne \mu_2$, then there exists a closed set F in $\mathscr{X}$ such that $\mu_1(F) \ne \mu_2(F)$. It is

not hard to show that for any $M > 0$ there exists a function $v = v_{M,F} \in \mathcal{U}(\mathcal{X})$ such that the right-hand side of (3.2) exceeds M. Since $M > 0$ is arbitrary, it follows that $J_{\pi_2}(\mu) = +\infty$. $\square$

We now assume that $\mu_1 = \mu_2$ and prove that $J_{\pi_2}(\mu) = I_{\mu,2}(\mu)$. Since $\mu_1 = \mu_2$,

$$
\begin{aligned}
(3.3) \quad J_{\pi_2}(\mu) = \sup_{u \in \mathcal{U}(\mathcal{X}^2)} \int_{\mathcal{X}} &\left[\int_{\mathcal{X}} \log u(x,y) \mu^*(x,dy) \right. \\
&\left. - \log \int_{\mathcal{X}} u(x,y) \pi(x,dy) \right] \mu_1(dx).
\end{aligned}
$$

We will use the following variational characterization of the relative entropy due to Donsker and Varadhan (1975).

LEMMA 3.2. *For ν and ρ Borel probability measures on $\mathcal{X}$,*

$$
I(\nu, \rho) = \sup_{v \in \mathcal{U}(\mathcal{X})} \left\{ \int_{\mathcal{X}} \log v(y) \nu(dy) - \log \int_{\mathcal{X}} v(y) \rho(dy) \right\}.
$$

According to this lemma, if $\mu_1 = \mu_2$, then

$$
J_{\pi_2}(\mu) \leq \int_{\mathcal{X}} \sup_{u \in \mathcal{U}(\mathcal{X}^2)} \left\{ \int_{\mathcal{X}} \log u(x,y) \mu^*(x,dy) - \log \int_{\mathcal{X}} u(x,y) \pi(x,dy) \right\} \mu_1(dx)
$$

$$
= \int_{\mathcal{X}} \sup_{v \in \mathcal{U}(\mathcal{X})} \left\{ \int_{\mathcal{X}} \log v(y) \mu^*(x,dy) - \log \int_{\mathcal{X}} v(y) \pi(x,dy) \right\} \mu_1(dx)
$$

$$
= \int_{\mathcal{X}} I(\mu^*(x,\cdot), \pi(x,\cdot)) \mu_1(dx)
$$

$$
= I_{\pi,2}(\mu).
$$

We now prove that if $\mu_1 = \mu_2$, then $J_{\pi_2}(\mu) \geq I_{\pi,2}(\mu)$. If $J_{\pi_2}(\mu) = +\infty$, then there is nothing to prove, so we assume that $J_{\pi_2}(\mu)$ is finite. For any $u \in \mathcal{U}(\mathcal{X}^2)$, (3.3) implies

$$
(3.4) \quad J_{\pi_2}(\mu) \geq \int_{\mathcal{X}} \left[\int_{\mathcal{X}} \log u(x,y) \mu^*(x,dy) - \log \int_{\mathcal{X}} u(x,y) \pi(x,dy) \right] \mu_1(dx).
$$

We define $\mathcal{V}(\mathcal{X}^2)$ to be the set of bounded measurable functions u on $\mathcal{X}^2$ such that $u \geq \varepsilon$ on $\mathcal{X}^2$ for some $\varepsilon = \varepsilon(u) > 0$. By tightness and Lusin's theorem, (3.4) continues to hold for all $u \in \mathcal{V}(\mathcal{X}^2)$. We need another lemma.

LEMMA 3.3. *Let μ be a Borel probability measure on $\mathcal{X}^2$ and define the set*

$$
A_\mu = \{ x \in \mathcal{X} : \mu^*(x,\cdot) \ll \pi(x,\cdot) \}.
$$

If $J_{\pi_2}(\mu)$ is finite, then $\mu_1(A_\mu) = 1$.

PROOF. For each $x \in \mathcal{X}$ define the measure $\phi(x, \cdot) = \mu^*(x, \cdot) + \pi(x, \cdot)$. Since $\pi(x, \cdot) \ll \phi(x, \cdot)$, we may write

$$k(x, y) = \frac{d\pi(x, \cdot)}{d\phi(x, \cdot)}(y), \qquad (x, y) \in \mathcal{X}^2.$$

According to a theorem of Doob [see Dellacherie and Meyer (1982), page 52], we may choose k to be a nonnegative measurable function of $(x, y) \in \mathcal{X}^2$. For any Borel set E in $\mathcal{X}$, we have

$$\pi(x, E) = \int_E k(x, y)\phi(x, dy) = \int_E k(x, y)\mu^*(x, dy) + \int_E k(x, y)\pi(x, dy)$$

or

(3.5) $$\int_E k(x, y)\mu^*(x, dy) = \int_E (1 - k(x, y))\pi(x, dy).$$

This implies that for each $x \in \mathcal{X}$, $\pi(x, B_x) = 0$ and $\pi(x, A_x) = 1$, where

$$B_x = \{y \in \mathcal{X} : k(x, y) = 0\}, \qquad A_x = \{y \in \mathcal{X} : k(x, y) > 0\}.$$

It follows from (3.5) that the measure $E \to \mu^*(x, B_x \cap E)$ is the singular part of $\mu^*(x, \cdot)$ relative to $\pi(x, \cdot)$. To prove the lemma, it suffices to show that

$$\int_{\mathcal{X}} \mu^*(x, B_x)\mu_1(dx) = 0.$$

For $\lambda > 0$ we define the function

$$u(x, y) = 1, \quad \text{for } x \in \mathcal{X}, \, y \in A_x,$$
$$= \lambda, \quad \text{for } x \in \mathcal{X}, \, y \in B_x.$$

Since $u(x, y) \in \mathcal{V}(\mathcal{X}^2)$, we have by (3.4) that

$$J_{\pi_2}(\mu) \geq (\log \lambda) \int_{\mathcal{X}} \mu^*(x, B_x)\mu_1(dx).$$

Since by hypothesis $J_{\pi_2}(\mu)$ is finite, it follows by taking $\lambda \to \infty$ that

$$\int_{\mathcal{X}} \mu^*(x, B_x)\mu_1(dx) = 0.$$

This completes the proof of the lemma. □

In order to complete the proof of Step 3, we need one more fact.

LEMMA 3.4. *We assume that $\pi(x, dy)$ satisfies Hypotheses 1.1(a)–(b). Then there exists a measure $\bar{\mu} \in \mathcal{M}(\mathcal{X}^2)$ such that*

(a) $J_{\pi_2}(\bar{\mu}) = 0$,
(b) *a regular conditional distribution $\bar{\mu}^*(x, dy)$ is given by $\pi(x, dy)$ and*
(c) *the one-dimensional marginal $\bar{\mu}_1$ is an invariant measure for π.*

PROOF.   Since π_2 satisfies Hypotheses 1.1(a)–(b) (Steps 1 and 2), Theorem 1.2 implies that the P_x-distributions of $M_{n,2}$ have a uniform large deviation property with the convex rate function J_{π_2}. It follows that $\inf\{J_{\pi_2}(\mu): \mu \in \mathcal{M}(\mathcal{X}^2)\} = 0$ and that there exists a measure $\bar{\mu} \in \mathcal{M}(\mathcal{X}^2)$ such that $J_{\pi_2}(\bar{\mu}) = 0$. This proves part (a). Parts (b) and (c) follow from the fact that $\bar{\mu}$ is an invariant measure for π_2 [Corollary 7.26 in Stroock (1984)]. □

We assume that $J_{\pi_2}(\mu)$ is finite and prove that

$$(3.6) \qquad J_{\pi_2}(\mu) \geq I_{\pi,2}(\mu) = \int_{\mathcal{X}} I(\mu^*(x, \cdot), \pi(x, \cdot))\mu_1(dx).$$

Let u be the function

$$(3.7) \qquad \begin{aligned} u(x, y) &= \frac{d\mu^*(x, \cdot)}{d\pi(x, \cdot)}(y), \quad \text{for } x \in A_\mu, \ y \in \mathcal{X}, \\ &= 1, \qquad\qquad\quad \text{for } x \in \mathcal{X} \setminus A_\mu, \ y \in \mathcal{X}, \end{aligned}$$

where $A_\mu = \{x \in \mathcal{X}: \mu^*(x, \cdot) \ll \pi(x, \cdot)\}$; according to Lemma 3.3, $\mu_1\{A_\mu\} = 1$. We prove (3.6) first under the assumption that $u \geq \theta$ on $\mathcal{X}^2$ for some $0 < \theta < 1$.

For n a positive integer, the function $u_n = u \wedge n$ is in $\mathcal{V}(\mathcal{X}^2)$, and so by (3.4) applied to u_n and monotone convergence ($u_n \uparrow u$),

$$\begin{aligned} J_{\pi_2}(\mu) &= J_{\pi_2}(\mu) + \int_{\mathcal{X}}\left[\log\int_{\mathcal{X}} u(x, y)\pi(x, dy)\right]\mu_1(dx) \\ &\geq \int_{\mathcal{X}}\left[\int_{\mathcal{X}} \log u(x, y)\mu^*(x, dy)\right]\mu_1(dx) \\ &= \int_{\mathcal{X}} I(\mu^*(x, \cdot), \pi(x, \cdot))\mu_1(dx) = I_{\pi,2}(\mu). \end{aligned}$$

We now remove the restriction that $u \geq \theta$ on $\mathcal{X}^2$. Let $\bar{\mu} \in \mathcal{M}(\mathcal{X}^2)$ be a measure satisfying the conclusions of Lemma 3.4. For $0 < \theta < 1$, define the measure

$$\mu_\theta = \theta\bar{\mu} + (1 - \theta)\mu \in \mathcal{M}(\mathcal{X}^2)$$

and the function

$$u_\theta(x, y) = \frac{d\mu_\theta^*(x, \cdot)}{d\pi(x, \cdot)}(y) = \theta + (1 - \theta)\frac{d\mu^*(x, \cdot)}{d\pi(x, \cdot)}(y),$$

$$\text{for } x \in A_\mu, \ y \in \mathcal{X},$$

$$= 1, \qquad\qquad \text{for } x \in X \setminus A_\mu, \ y \in \mathcal{X}.$$

Thus $u_\theta = \theta + (1 - \theta)u$, where u is defined in (3.7). Since $u_\theta \geq \theta$ on $\mathcal{X}^2$, we may apply the previous argument to the measure μ_θ, obtaining

$$\begin{aligned} J_{\pi_2}(\mu_\theta) &= J_{\pi_2}(\theta\bar{\mu} + (1 - \theta)\mu) \\ &\geq I_{\pi,2}(\mu_\theta) \\ &= \int_{\mathcal{X}} I(\mu_\theta^*(x, \cdot), \pi(x, \cdot))[\theta\bar{\mu}_1(dx) + (1 - \theta)\mu_1(dx)]. \end{aligned}$$

Since J_{π_2} is convex, $J_{\pi_2}(\bar{\mu}) = 0$, and $I(\mu_\theta^*(x, \cdot), \pi(x, \cdot)) \geq 0$, we see that

$$J_{\pi_2}(\mu) \geq \int_{\mathscr{X}} I(\mu_\theta^*(x, \cdot), \pi(x, \cdot))\mu_1(dx)$$

$$= \int_{\mathscr{X}} \left[\int_{\mathscr{X}} u_\theta(x, y)\log u_\theta(x, y)\pi(x, dy)\right]\mu_1(dx).$$

$$= \int_{\mathscr{X}} \left[\theta \int_{\mathscr{X}} \log u_\theta(x, y)\pi(x, dy)\right.$$

$$\left. + (1 - \theta)\int_{\mathscr{X}} u(x, y)\log u_\theta(x, y)\pi(x, dy)\right]\mu_1(dx)$$

$$\geq \theta \log \theta + (1 - \theta)^2 \int_{\mathscr{X}} \left[\int_{\mathscr{X}} u(x, y)\log u(x, y)\pi(x, dy)\right]\mu_1(dx).$$

The last inequality uses the fact that $t \to \log t$ is concave. Take $\theta \to 0$ and conclude

$$J_{\pi_2}(\mu) \geq \int_{\mathscr{X}} \left[\int_{\mathscr{X}} u(x, y)\log u(x, y)\pi(x, dy)\right]\mu_1(dx) = I_{\pi, 2}(\mu).$$

In Lemma 3.1, we proved that $J_{\pi_2}(\mu) = +\infty$ if $\mu_1 \neq \mu_2$. Thus if $\mu_1 \neq \mu_2$, then $J_{\pi_2}(\mu) = I_{\pi, 2}(\mu) = +\infty$. Assuming $\mu_1 = \mu_2$, we have shown also that $J_{\pi_2}(\mu) = I_{\pi, 2}(\mu)$. This completes the proof of Step 3. The proof of Theorem 1.4 is complete. $\square$

REMARK 3.5. The purpose of this remark is to point out why the conclusions of Theorem 1.4 remain valid under the weaker hypotheses given in Remark 1.3(a)–(b). As we have seen, Theorem 1.4 is proved in two parts. Part 1 shows that the P_x-distributions of $M_{n, \alpha}(\omega, \cdot)$ have a uniform large deviation property with the convex rate function $J_{\pi_\alpha}(\mu)$. Part 2 identifies $J_{\pi_\alpha}(\mu)$ with the function $I_{\pi, \alpha}(\mu)$ and with the relative entropy $I(\mu, \mu_1 \otimes \pi)$. Part 1 is proved by checking that the transition probability function $\pi_\alpha(x, dy)$ satisfies the hypotheses of Theorem 1.2 [Hypotheses 1.1(a)–(b)]. Since the conclusions of Theorem 1.2 remain valid under the weaker hypotheses given in Remark 1.3(a)–(b), Part 1 of the proof of Theorem 1.4 also remains valid under these weaker hypotheses. Part 2 of the proof of Theorem 1.4 is then proved exactly as in the remainder of Section 3 [Lemma 3.1 to the end].

REFERENCES

CSISZÁR, I., COVER T. M. and CHOI, B. S. (1987). Conditional limit theorems under Markov conditioning. *IEEE Trans. Inform. Theory* **IT-33** 788–801.

DELLACHERIE, C. and MEYER, P.-A. (1982). *Probabilities and Potential B*. North-Holland, Amsterdam.

DONSKER, M. D. and VARADHAN, S. R. S. (1975). 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. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461.

ELLIS, R. S. (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York.

ELLIS, R. S. and WYNER, A. D. (1987). Uniform large deviation property of the empirical process of a Markov chain. Preprint, Univ. Massachusetts and AT & T Bell Laboratories.

NATARAJAN, S. (1985). Large deviations, hypotheses testing, and source coding for finite Markov chains. *IEEE Trans. Inform. Theory* **IT-31** 360–365.

NATARAJAN, S. (1986). Uniform large deviation property of empirical transition counts of a finite Markov chain. Preprint, Indian Statist. Inst.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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

STROOCK, D. W. (1987). Private communication based on a new edition of Stroock (1984) (new edition to appear).

SZULGA, J., WOYCZYNSKI, W. A., MANN, J. A. and TJATJOPOULOS, G. J. (1987). One dimensional monolayers: Randomized partition functions and large deviation techniques. Preprint, Case Western Reserve Univ.

DEPARTMENT OF MATHEMATICS AND STATISTICS
LEDERLE GRADUATE RESEARCH CENTER TOWERS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003