Chapter IX

Level-3 Large Deviations for I.I.D. Random Vectors

IX.1. Statement of Results

Theorem II.4.4 stated the level-3 large deviation property for i.i.d. random vectors taking values in $\mathbb{R}^d$. In this chapter, we prove Theorem II.4.4 in the special case of i.i.d. random variables with a finite state space. This version of the theorem covers the applications of level-3 large deviations which were made in Chapters III, IV, and V to the Gibbs variational principle. Theorem II.4.4 can also be proved via the methods of Donsker and Varadhan (1983a). The main result in that paper is a level-3 theorem for continuous parameter Markov processes taking values in a complete separable metric space.[1]

Let ρ be a Borel probability measure on $\mathbb{R}$ whose support is a finite set Γ. We topologize Γ by the discrete topology and $\Gamma^{\mathbb{Z}}$ by the product topology. With respect to the probability space $(\Gamma^{\mathbb{Z}}, \mathscr{B}(\Gamma^{\mathbb{Z}}), P_\rho)$, the coordinate representation process $X_j(\omega) = \omega_j$ is a sequence of i.i.d. random variables distributed by ρ. $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ denotes the set of strictly stationary probability measures on $\mathscr{B}(\Gamma^{\mathbb{Z}})$ with the topology of weak convergence. The empirical process is defined as

$$R_n(\omega, \cdot) = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k X(n,\omega)}(\cdot), \qquad n = 1, 2, \ldots, \quad \omega \in \Gamma^{\mathbb{Z}},$$

where T is the shift mapping on $\Gamma^{\mathbb{Z}}$ and $X(n, \omega)$ is the periodic point in $\Gamma^{\mathbb{Z}}$ obtained by repeating $(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))$ periodically. For each Borel subset B of $\Gamma^{\mathbb{Z}}$, $R_n(\omega, B)$ is the relative frequency with which $X(n, \omega)$, $TX(n, \omega), \ldots, T^{n-1}X(n, \omega)$ is in B. Thus $R_n(\omega, \cdot)$ is for each ω an element of $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. For $P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$, $\tilde{P}_\omega$ denotes a regular conditional distribution, with respect to P, of X_1 given the σ-field $\mathscr{F}\{X_j; j \le 0\}$. The level-3 entropy function is defined as

$$I_\rho^{(3)}(P) = \int_{\Gamma^{\mathbb{Z}}} I_\rho^{(2)}(\tilde{P}_\omega) P(d\omega), \tag{9.1}$$

where $I_\rho^{(2)}(\tilde{P}_\omega)$ is the relative entropy of $\tilde{P}_\omega$ with respect to ρ.

The following theorem is Theorem II.4.4 for the case of a finite state space.

Theorem IX.1.1. *Let ρ be a Borel probability measure on $\mathbb{R}$ whose support is a finite set Γ. Then the following conclusions hold.*

(a) *$\{Q_n^{(3)}\}$, the P_ρ-distributions on $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$ of the empirical processes $\{R_n\}$, have a large deviation property with $a_n = n$ and entropy function $I_\rho^{(3)}$.*

(b) *$I_\rho^{(3)}(P)$ is an affine function of P. $I_\rho^{(3)}(P)$ measures the discrepancy between P and the infinite product measure P_ρ in the sense that $I_\rho^{(3)}(P) \geq 0$ with equality if and only if $P = P_\rho$.*

If A is a nonempty subset of $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$, then $I_\rho^{(3)}(A)$ denotes the infimum of $I_\rho^{(3)}$ over A. $I_\rho^{(3)}(\phi)$ equals ∞. In order to prove part (a) of the theorem, we must verify the following hypotheses.

(i) $I_\rho^{(3)}(P)$ is lower semicontinuous.
(ii) $I_\rho^{(3)}(P)$ has compact level sets.
(iii) $\limsup_{n\to\infty} n^{-1} \log Q_n^{(3)}\{K\} \leq -I_\rho^{(3)}(K)$ for each closed set K in $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$.
(iv) $\liminf_{n\to\infty} n^{-1} \log Q_n^{(3)}\{G\} \geq -I_\rho^{(3)}(G)$ for each open set G in $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$.

Hypotheses (i) and (ii) and part (b) of the theorem will be proved in Section IX.2. We prove hypotheses (iii) and (iv) by first showing the large deviation bounds for finite-dimensional sets in $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$. The proof of the bounds for such sets depends on the following facts.

(v) For each $\alpha \geq 1$ the distributions of the α-dimensional marginals of $\{R_n\}$ have a large deviation property with entropy function denoted by $I_{\rho,\alpha}^{(3)}$.
(vi) $I_{\rho,\alpha}^{(3)}$ is related to $I_\rho^{(3)}$ by the contraction principle

$$\inf\{I_\rho^{(3)}(P) \colon P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}}), \pi_\alpha P = \tau\} = I_{\rho,\alpha}^{(3)}(\tau),$$

where $\tau = \pi_\alpha P$ is the fixed α-dimensional marginal of P.

Item (vi) is proved in Section IX.3; (v), (iii), and (iv) are proved in Section IX.4.

IX.2. Properties of the Level-3 Entropy Function

Let ρ be a Borel probability measure on $\mathbb{R}$ whose support is a finite set $\Gamma = \{x_1, x_2, \ldots, x_r\}$ with $x_1 < x_2 < \cdots < x_r$. Set $\rho_i = \rho\{x_i\} > 0$. Let α be a positive integer and π_α the projection of $\Gamma^{\mathbb{Z}}$ onto Γ^α defined by $\pi_\alpha \omega = (\omega_1, \ldots, \omega_\alpha)$. If P is a strictly stationary probability measure on $\mathscr{B}(\Gamma^{\mathbb{Z}})$, then define a probability measure $\pi_\alpha P$ on $\mathscr{B}(\Gamma^\alpha)$ by requiring

$$\pi_\alpha P\{F\} = P\{\pi_\alpha^{-1} F\} = P\{\omega \in \Gamma^{\mathbb{Z}} \colon (\omega_1, \ldots, \omega_\alpha) \in F\}$$

for subsets F of Γ^α. The measure $\pi_\alpha P$ is called the *α-dimensional marginal* of P. We consider the quantity

$$I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P) = \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P\{\omega\} \log \frac{\pi_\alpha P\{\omega\}}{\pi_\alpha P_\rho\{\omega\}},$$

which is the relative entropy of $\pi_\alpha P$ with respect to $\pi_\alpha P_\rho$. In Chapter I, the level-3 entropy function $I_\rho^{(3)}(P)$ was defined as $\lim_{\alpha\to\infty} \alpha^{-1} I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P)$ [see (1.37)]. We will show in Theorem IX.2.3 that this limit exists and coincides with the quantity $I_\rho^{(3)}(P)$ defined in (9.1).

Given elements $x_{i_1}, \ldots, x_{i_\alpha}$ in Γ, set

$$p(x_{i_1}, \ldots, x_{i_\alpha}) = P\{X_1 = x_{i_1}, \ldots, X_\alpha = x_{i_\alpha}\}$$

and provided $P\{X_1 = x_{i_1}, \ldots, X_{\alpha-1} = x_{i_{\alpha-1}}\} > 0$, let $p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}})$ denote the conditional probability $P\{X_\alpha = x_{i_\alpha}|X_1 = x_{i_1}, \ldots, X_{\alpha-1} = x_{i_{\alpha-1}}\}$. We define

$$H_{\rho,\alpha}(P) = \begin{cases} \sum_{i_1=1}^{r} p(x_{i_1}) \log[p(x_{i_1})/\rho_{i_1}] & \text{for } \alpha = 1, \\ \sum p(x_{i_1}, \ldots, x_{i_\alpha}) \log[p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}})/\rho_{i_\alpha}] & \text{for } \alpha \geq 2. \end{cases}$$

For $\alpha \geq 2$ the sum defining $H_{\rho,\alpha}(P)$ runs over all $i_1, \ldots, i_{\alpha-1}, i_\alpha$ for which $p(x_{i_1}, \ldots, x_{i_{\alpha-1}}) > 0$. Since $0 \log 0 = 0$, the sum may be restricted to all $i_1, \ldots, i_\alpha$ for which $p(x_{i_1}, \ldots, x_{i_\alpha}) > 0$. The quantity $-\sum p(x_{i_1}, \ldots, x_{i_\alpha}) \cdot \log p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}})$ is known as the *conditional entropy of* X_α *given* $X_1, \ldots, X_{\alpha-1}$.

Lemma IX.2.1 (a) *Let* $p_1, \ldots, p_N$ *and* $q_1, \ldots, q_N$ *be non-negative real numbers such that* $\sum_{i=1}^N p_i = 1$, $\sum_{i=1}^N q_i = 1$, *and* $p_i = 0$ *whenever* $q_i = 0$. *Then*

$$\sum_{i=1}^N p_i \log p_i \geq \sum_{i=1}^N p_i \log q_i$$

with equality if and only if $p_i = q_i$ *for all* i.

(b) $I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P) = H_{\rho,1}(P) + H_{\rho,2}(P) + \cdots + H_{\rho,\alpha-1}(P) + H_{\rho,\alpha}(P)$.

(c) $0 \leq H_{\rho,1}(P) \leq H_{\rho,2}(P) \leq \cdots \leq H_{\rho,\alpha-1}(P) \leq H_{\rho,\alpha}(P)$.

(d) $H_{\rho,2}(P) = H_{\rho,1}(P)$ *if and only if* X_1 *and* X_2 *are independent.*

(e) *For* $\alpha \geq 3$, $H_{\rho,\alpha}(P) = H_{\rho,\alpha-1}(P)$ *if and only if* X_1 *and* X_α *are conditionally independent given* $X_2, \ldots, X_{\alpha-1}$; *that is, if and only if*

$$p(x_{i_\alpha}|x_{i_1}, x_{i_2}, \ldots, x_{i_{\alpha-1}}) = p(x_{i_\alpha}|x_{i_2}, \ldots, x_{i_{\alpha-1}})$$

$$\text{whenever } p(x_{i_1}, x_{i_2}, \ldots, x_{i_{\alpha-1}}) > 0.*$$

Proof. (a) We have

$$\sum_{i=1}^N p_i \log p_i - \sum_{i=1}^N p_i \log q_i = \sum_{q_i>0} p_i \log \frac{p_i}{q_i}.$$

The latter is non-negative and equals 0 if and only if $p_i = q_i$ whenever $q_i > 0$ [Proposition I.4.1(b)]. Since by hypothesis $p_i = q_i$ whenever $q_i = 0$, the proof of part (a) is done.

* See page 301.

(b) If $\alpha \geq 2$ and $p(x_{i_1}, \ldots, x_{i_\alpha}) > 0$, then

$$p(x_{i_1}, \ldots, x_{i_\alpha}) = p(x_{i_1}) \cdot \prod_{\beta=2}^{\alpha} p(x_{i_\beta} | x_{i_1}, \ldots, x_{i_{\beta-1}}).$$

Hence for $\alpha \geq 2$

$$I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P) = \sum p(x_{i_1}, \ldots, x_{i_\alpha}) \left[\log \frac{p(x_{i_1})}{\rho_{i_1}} + \sum_{\beta=2}^{\alpha} \log \frac{p(x_{i_\beta} | x_{i_1}, \ldots, x_{i_{\beta-1}})}{\rho_{i_\beta}} \right]$$

$$= \sum_{\beta=1}^{\alpha} H_{\rho,\beta}(P).$$

(c)–(e) $H_{\rho,1}(P)$ equals the relative entropy of $\pi_1 P$ with respect to ρ and so is non-negative. Since P is strictly stationary,

$$H_{\rho,1}(P) = \sum_{i_1, i_2 = 1}^{r} p(x_{i_1}, x_{i_2}) \log \frac{p(x_{i_2})}{\rho_{i_2}}.$$

The sum may be restricted to i_1, i_2 for which $p(x_{i_1}) > 0$. For if $p(x_{i_1}) = 0$, then $p(x_{i_1}, x_{i_2}) = 0$ for all i_2, and there is no contribution to the sum. Thus

$$H_{\rho,2}(P) - H_{\rho,1}(P) = \sum_{p(x_{i_1})>0} p(x_{i_1}) \left\{ \sum_{i_2=1}^{r} p(x_{i_2}|x_{i_1}) \log p(x_{i_2}|x_{i_1}) - \sum_{i_2=1}^{r} p(x_{i_2}|x_{i_1}) \log p(x_{i_2}) \right\}.$$

By part (a), $H_{\rho,2}(P) \geq H_{\rho,1}(P)$ with equality if and only if $p(x_{i_2}|x_{i_1}) = p(x_{i_2})$ whenever $p(x_{i_1}) > 0$. The latter condition is equivalent to the independence of X_1 and X_2. Similarly, for $\alpha \geq 3$,

$$H_{\rho,\alpha}(P) - H_{\rho,\alpha-1}(P)$$

$$= \sum p(x_{i_1}, \ldots, x_{i_{\alpha-1}}) \left\{ \sum_{i_\alpha=1}^{r} p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}}) \log p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}}) - \sum_{i_\alpha=1}^{r} p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}}) \log p(x_{i_\alpha}|x_{i_2}, \ldots, x_{i_{\alpha-1}}) \right\},$$

where the outer sum runs over all $i_1, \ldots, i_{\alpha-1}$ for which $p(x_{i_1}, \ldots, x_{i_{\alpha-1}}) > 0$. By part (a), $H_{\rho,\alpha}(P) \geq H_{\rho,\alpha-1}(P)$ with equality if and only if

$$p(x_{i_\alpha}|x_{i_1}, \ldots, x_{i_{\alpha-1}}) = p(x_{i_\alpha}|x_{i_2}, \ldots, x_{i_{\alpha-1}}) \quad \text{whenever } p(x_{i_1}, \ldots, x_{i_{\alpha-1}}) > 0. \qquad \square$$

We also need the following lemma, which arises in the proof of the Shannon–McMillan–Breiman theorem in information theory [Billingsley (1965, page 131)].

Lemma IX.2.2. *For each* $i_1 \in \{1, \ldots, r\}$,

$$0 \le \int_{\{X_1 = x_{i_1}\}} \sup_{n \ge 0} [-\log P\{X_1 = x_{i_1} | X_{-n}, \ldots, X_0\}(\omega)] P(d\omega) < \infty.$$

Proof. For n a non-negative integer, $i \in \{1, \ldots, r\}$, and $\omega \in \Gamma^{\mathbb{Z}}$, define

$$f_n^{(i)}(\omega) = -\log P\{X_1 = x_i | X_{-n}, \ldots, X_0\}(\omega),$$

$$g_n(\omega) = \sum_{i=1}^{r} f_n^{(i)}(\omega) \cdot \chi_{\{X_1 = x_i\}}(\omega).$$

Since $0 \le P\{X_1 = x_i | X_{-n}, \ldots, X_0\}(\omega) \le 1$ P-a.s., it suffices to prove that

$$\int_{\Gamma^{\mathbb{Z}}} \sup_{n \ge 0} g_n(\omega) P(d\omega) < \infty. \tag{9.2}$$

If $A_n = \{\omega \in \Gamma^{\mathbb{Z}} : \max_{0 \le j < n} g_j(\omega) \le \lambda < g_n(\omega)\}$, $\lambda \ge 0$, then

$$P\{A_n\} = \sum_{i=1}^{r} P\{\{X_1 = x_i\} \cap A_n\} = \sum_{i=1}^{r} P\{\{X_1 = x_i\} \cap B_n^{(i)}\},$$

where $B_n^{(i)} = \{\omega \in \Gamma^{\mathbb{Z}} : \max_{0 \le j < n} f_j^{(i)}(\omega) \le \lambda < f_n^{(i)}(\omega)\}$. $B_n^{(i)}$ is in the σ-field generated by $X_{-n}, \ldots, X_0$, and so

$$P\{\{X_1 = x_i\} \cap B_n^{(i)}\} = \int_{B_n^{(i)}} P\{X_1 = x_i | X_{-n}, \ldots, X_0\}(\omega) P(d\omega)$$

$$= \int_{B_n^{(i)}} \exp(-f_n^{(i)}(\omega)) P(d\omega) \le e^{-\lambda} P\{B_n^{(i)}\}.$$

Since the sets $\{B_n^{(i)}; n = 0, 1, \ldots\}$ are disjoint,

$$\sum_{n=0}^{\infty} P\{A_n\} \le e^{-\lambda} \sum_{i=1}^{r} \sum_{n=0}^{\infty} P\{B_n^{(i)}\} \le r e^{-\lambda}.$$

Hence $P\{\omega \in \Gamma^{\mathbb{Z}} : \sup_{n \ge 0} g_n(\omega) > \lambda\} \le r e^{-\lambda}$, and (9.2) follows. □

Theorem IX.2.3. *For* $P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}})$, *define* $I_\rho^{(3)}(P)$ *by* (9.1). *Then* $\lim_{\alpha \to \infty} \alpha^{-1} I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P)$ *exists*, $\lim_{\alpha \to \infty} H_{\rho,\alpha}(P)$ *exists, and*

$$\lim_{\alpha \to \infty} \frac{1}{\alpha} I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P) = \lim_{\alpha \to \infty} H_{\rho,\alpha}(P) = I_\rho^{(3)}(P) < \infty.\text{*}$$

Proof. By Lemma IX.2.1(b), it suffices to prove that $\lim_{\alpha \to \infty} H_{\rho,\alpha}(P)$ exists and equals $I_\rho^{(3)}(P)$. Since P is strictly stationary, we have for $n \ge 0$

$$H_{\rho,n+2}(P) = \sum_{i_1=1}^{r} \int_{\{X_1 = x_{i_1}\}} \log[P\{X_1 = x_{i_1} | X_{-n}, \ldots, X_0\}/\rho_{i_1}] dP.$$

*There is an elementary proof of the existence of $\lim_{\alpha \to \infty} \alpha^{-1} I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P)$ which does not identify the form of the limit. See Note 2.

As $n \to \infty$, $P\{X_1 = x_{i_1} | X_{-n}, \ldots, X_0\}$ converges to $P\{X_1 = x_{i_1} | \{X_j; j \le 0\}\}$ almost surely [Theorem A.6.2(a)]. By the Lebesgue dominated convergence theorem and Lemma IX.2.2, we conclude that $\lim_{\alpha\to\infty} H_{\rho,\alpha}(P)$ exists and*

$$\begin{aligned}
\lim_{\alpha\to\infty} H_{\rho,\alpha}(P) &= \sum_{i_1=1}^{r} \int_{\{X_1 = x_{i_1}\}} \log[P\{X_1 = x_{i_1} | \{X_j; j \le 0\}\}/\rho_{i_1}]dP \\
&= \sum_{i_1=1}^{r} \int_{\Gamma^{\mathbb{Z}}} \log[P\{X_1 = x_{i_1} | \{X_j; j \le 0\}\}/\rho_{i_1}] \\
&\quad \cdot P\{X_1 = x_{i_1} | \{X_j; j \le 0\}\} dP \\
&= \int_{\Gamma^{\mathbb{Z}}} \int_{\Gamma} \log \frac{\tilde{P}_\omega\{x\}}{\rho\{x\}} \tilde{P}_\omega(dx) P(d\omega) \\
&= \int_{\Gamma^{\mathbb{Z}}} I_\rho^{(2)}(\tilde{P}_\omega) P(d\omega) = I_\rho^{(3)}(P) < \infty. \qquad \square
\end{aligned}$$

In order to prove properties of $I_\rho^{(3)}(P)$, we need a standard lemma.

Lemma IX.2.4. *Let $b_1, b_2, \ldots$ be a superadditive sequence of real numbers; i.e., $b_{m+n} \ge b_m + b_n$ for all positive integers m and n. Then $\lim_{n\to\infty} b_n/n = \sup_{n\ge1} b_n/n$.*

Proof. Define $s = \sup_{n\ge1} b_n/n$ and suppose that s is not ∞. Given $\varepsilon > 0$ choose a positive integer k such that $b_k/k > s - \varepsilon$. Any positive integer n has the form $n = mk + l$, where $m \ge 0$ and $0 \le l < k$. By hypothesis

$$b_n = b_{mk+l} \ge mb_k + lb_1.$$

Hence $s \ge \liminf_{n\to\infty} b_n/n \ge \liminf_{n\to\infty} mb_k/n = b_k/k > s - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, it follows that $\lim_{n\to\infty} b_n/n = s$ if s is finite. The proof is similar if s is ∞. $\square$

$I_\rho^{(3)}(P)$ is lower semicontinuous. We first prove that for $P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}})$ the sequence of relative entropies $I_\alpha = I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P)$ is superadditive. Let α and β be arbitrary positive integers and introduce the notation

$$\omega \circ \bar{\omega} = (\omega_1, \omega_2, \ldots, \omega_\alpha, \bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_\beta) \in \Gamma^{\alpha+\beta} \qquad \text{for } \omega \in \Gamma^\alpha, \bar{\omega} \in \Gamma^\beta.$$

Since $\pi_{\alpha+\beta} P_\rho\{\omega \circ \bar{\omega}\} = \pi_\alpha P_\rho\{\omega\} \cdot \pi_\beta P_\rho\{\bar{\omega}\}$,

$$I_{\alpha+\beta} = \sum_{\omega\in\Gamma^\alpha} \sum_{\bar{\omega}\in\Gamma^\beta} \pi_{\alpha+\beta} P\{\omega \circ \bar{\omega}\} \log \frac{\pi_{\alpha+\beta} P\{\omega \circ \bar{\omega}\}}{\pi_\alpha P_\rho\{\omega\} \cdot \pi_\beta P_\rho\{\bar{\omega}\}}.$$

Since P is strictly stationary, $\sum_{\omega\in\Gamma^\alpha} \pi_{\alpha+\beta} P\{\omega \circ \bar{\omega}\} = \pi_\beta P\{\bar{\omega}\}$, and so

*The second equality in the display uses (A.3) [page 300].

$$I_{\alpha+\beta} - I_\alpha - I_\beta = \sum_{\omega \in \Gamma^\alpha} \sum_{\bar{\omega} \in \Gamma^\beta} \pi_{\alpha+\beta} P\{\omega \circ \bar{\omega}\} \log \pi_{\alpha+\beta} P\{\omega \circ \bar{\omega}\}$$
$$- \sum_{\omega \in \Gamma^\alpha} \sum_{\bar{\omega} \in \Gamma^\beta} \pi_{\alpha+\beta} P\{\omega \circ \bar{\omega}\} \log[\pi_\alpha P\{\omega\} \cdot \pi_\beta P\{\bar{\omega}\}].$$

By Lemma IX.2.1(a) $I_{\alpha+\beta} - I_\alpha - I_\beta \geq 0$. It follows from Lemma IX.2.4 that

$$I_\rho^{(3)}(P) = \lim_{\alpha \to \infty} \frac{1}{\alpha} I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P) = \sup_{\alpha \geq 1} \frac{1}{\alpha} I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P).$$

Since the mapping $P \to \pi_\alpha P$ is continuous, each function $I^{(2)}_{\pi_\alpha P_\rho}(\pi_\alpha P)$ is a continuous function of $P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}})$. As the supremum of a sequence of continuous functions, $I_\rho^{(3)}(P)$ is a lower semicontinuous function of $P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}})$.

$I_\rho^{(3)}(P)$ *has compact level sets.* $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$ is a compact metric space [Theorem A.9.2(c)]. Since $I_\rho^{(3)}(P)$ is lower semicontinuous, its level sets are closed subsets of $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$ and thus are compact subsets of $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$.

$I_\rho^{(3)}(P)$ *is affine.* We show that if P and Q are measures in $\mathcal{M}_s(\Gamma^{\mathbb{Z}})$, then for every $0 < \lambda < 1$

$$(9.3) \qquad I_\rho^{(3)}(\lambda P + (1 - \lambda)Q) = \lambda I_\rho^{(3)}(P) + (1 - \lambda) I_\rho^{(3)}(Q).$$

Write $p(\omega)$ for $\pi_\alpha P\{\omega\}/\pi_\alpha P_\rho\{\omega\}$ and $q\{\omega\}$ for $\pi_\alpha Q\{\omega\}/\pi_\alpha P_\rho\{\omega\}$, $\omega \in \Gamma^\alpha$. Since $x \log x$, $x \geq 0$, is convex and $\log x$, $x > 0$, is nondecreasing,

$$\lambda I^{(2)}_{\pi_\alpha P_\rho}(P) + (1 - \lambda) I^{(2)}_{\pi_\alpha P_\rho}(Q) = \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P_\rho\{\omega\} [\lambda p(\omega) \log p(\omega)$$
$$+ (1 - \lambda) q(\omega) \log q(\omega)]$$
$$\geq \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P_\rho\{\omega\} [\lambda p(\omega) + (1 - \lambda) q(\omega)]$$
$$\cdot \log[\lambda p(\omega) + (1 - \lambda) q(\omega)]$$
$$\geq \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P_\rho\{\omega\} [\lambda p(\omega) \log(\lambda p(\omega))$$
$$+ (1 - \lambda) q(\omega) \log((1 - \lambda) q(\omega))]$$
$$= \lambda \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P_\rho\{\omega\} \cdot p(\omega) \log p(\omega)$$
$$+ (1 - \lambda) \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P_\rho\{\omega\} \cdot q(\omega) \log q(\omega)$$
$$+ \lambda \log \lambda + (1 - \lambda) \log(1 - \lambda)$$
$$\geq \lambda I^{(2)}_{\pi_\alpha P_\rho}(P) + (1 - \lambda) I^{(2)}_{\pi_\alpha P_\rho}(Q) - \log 2.$$

The sum on the third and fourth lines is $I^{(2)}_{\pi_\alpha P_\rho}(\lambda P + (1 - \lambda)Q)$. Dividing each term in the display by α and letting α tend to ∞, we obtain (9.3).

$I_\rho^{(3)}(P) \geq 0$ *with equality if and only if* $P = P_\rho$. Given $P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}})$, let ν equal $\pi_1 P$. If $\nu = \rho$, then by Theorem IX.3.1 below

$$I_\rho^{(3)}(P) \geq I_\rho^{(2)}(\rho) = 0 \qquad \text{with equality if and only if } P = P_\rho.$$

If $\nu \neq \rho$, then by Theorems IX.3.1 and VIII.2.1(b)

$$I_\rho^{(3)}(P) \geq I_\rho^{(2)}(\nu) > I_\rho^{(2)}(\rho) = 0.$$

This completes the proofs of properties of $I_\rho^{(3)}$.

IX.3. Contraction Principles

The first theorem is a contraction principle relating levels-2 and 3.

Theorem IX.3.1. *Let ν be a probability measure on $\mathscr{B}(\Gamma)$ and P_ν the infinite product measure on $\mathscr{B}(\Gamma^{\mathbb{Z}})$ with $\pi_1 P = \nu$. Then $I_\rho^{(3)}(P)$ attains its infimum over the set $\{P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}): \pi_1 P = \nu\}$ at the unique measure P_ν and*

$$\inf\{I_\rho^{(3)}(P): P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \pi_1 P = \nu\} = I_\rho^{(3)}(P_\nu) = I_\rho^{(2)}(\nu).$$

Proof. Let $\mathscr{A}$ be the set $\{P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}): \pi_1 P = \nu\}$ and P any measure in $\mathscr{A}$. $H_{\rho,1}(P)$ equals $\sum_{i=1}^r \nu_i \log(\nu_i/\rho_i) = I_\rho^{(2)}(\nu)$ $(\nu_i = \nu\{x_i\})$. By Lemma IX.2.1(c) and Theorem IX.2.3, for any $k \geq 2$

$$(9.4) \qquad 0 \leq H_{\rho,1}(P) = I_\rho^{(2)}(\nu) \leq \cdots \leq H_{\rho,k}(P) \leq H_{\rho,k+1}(P) \uparrow I_\rho^{(3)}(P).$$

If $P = P_\nu$, then for any $k \geq 1$ $H_{\rho,k}(P_\nu) = I_\rho^{(2)}(\nu) = I_\rho^{(3)}(P_\nu)$. We see that $I_\rho^{(3)}(P)$ attains its infimum over $\mathscr{A}$ at the measure $P = P_\nu$ and that the infimum equals $I_\rho^{(2)}(\nu)$. The proof is done once we show that $P = P_\nu$ is the unique measure in $\mathscr{A}$ for which $I_\rho^{(3)}(P)$ equals $I_\rho^{(2)}(\nu)$.

Suppose that $I_\rho^{(3)}(P)$ equals $I_\rho^{(2)}(\nu)$ for some P in $\mathscr{A}$. We prove by induction that for all $k \geq 1$

$$(9.5) \qquad p(x_{i_1}, \ldots, x_{i_k}) = \nu_{i_1} \ldots \nu_{i_k}.$$

This will imply that P equals P_ν. Formula (9.5) holds for $k = 1$ since $\pi_1 P$ equals ν. Since $I_\rho^{(3)}(P) = I_\rho^{(2)}(\nu)$, (9.4) shows that $H_{\rho,k}(P)$ equals $H_{\rho,1}(P)$ for all $k \geq 2$. With respect to P, X_1 and X_2 are independent [Lemma IX.2.1(d)], and so (9.5) holds for $k = 2$. Assume that (9.5) has been shown for $k = 1$, $2, \ldots, c-1$, some $c \geq 3$. With respect to P, X_1 and X_c are conditionally independent given $X_2, \ldots, X_{c-1}$ [Lemma IX.2.1(e)]. If $p(x_{i_1}, \ldots, x_{i_{c-1}}) > 0$, then by the induction hypothesis and strict stationarity

$$p(x_{i_1}, \ldots, x_{i_c}) = p(x_{i_1}, \ldots, x_{i_{c-1}}) p(x_{i_c} | x_{i_2}, \ldots, x_{i_{c-1}}) = \nu_{i_1} \ldots \nu_{i_{c-1}} \nu_{i_c}.$$

If $p(x_{i_1}, \ldots, x_{i_{c-1}}) = 0$, then $p(x_{i_1}, \ldots, x_{i_c}) = 0 = \nu_{i_1} \cdots \nu_{i_c}$. Thus (9.5) holds for $k = c$. The proof of the theorem is complete. □

We now generalize the contraction principle just proved by calculating the infimum of $I_\rho^{(3)}(P)$ over all measures $P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$ with fixed marginal $\tau = \pi_\alpha P$, $\alpha \in \{2, 3, \ldots\}$. Denote by $\mathscr{M}_s(\Gamma^\alpha)$ the set of probability measures τ

on $\mathscr{B}(\Gamma^\alpha)$ which have the form $\tau = \pi_\alpha P$ for some $P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$. For $\tau \in \mathscr{M}_s(\Gamma^\alpha)$, set $\tau_{i_1 \cdots i_\alpha} = \tau\{x_{i_1}, \ldots, x_{i_\alpha}\}$ and $(v_\tau)_{i_1 \cdots i_{\alpha-1}} = \sum_{i_\alpha=1}^r \tau_{i_1 \cdots i_\alpha}$ and define

$$I_{\rho,\alpha}^{(3)}(\tau) = \sum \tau_{i_1 \cdots i_\alpha} \log \frac{\tau_{i_1 \cdots i_\alpha}}{(v_\tau)_{i_1 \cdots i_{\alpha-1}} \rho_{i_\alpha}}, \tag{9.6}$$

where the sum runs over all $i_1, \ldots, i_{\alpha-1}, i_\alpha$ for which $(v_\tau)_{i_1 \cdots i_{\alpha-1}} > 0$. $I_{\rho,\alpha}^{(3)}(\tau)$ is well defined ($0 \log 0 = 0$) and equals the relative entropy of τ with respect to $\{(v_\tau)_{i_1 \cdots i_{\alpha-1}} \rho_{i_\alpha}\}$. We have encountered the function $I_{\rho,2}^{(3)}$ as the entropy function in Theorem I.5.1, which was a large deviation result for the quantities $\{M_n(\omega, \cdot)\}$ (empirical pair measures). The latter measures are related to the empirical processes by the formula $M_n(\omega, \cdot) = \pi_2 R_n(\omega, \cdot)$ [see (1.35)].

In the next section, the distributions of $\{\pi_\alpha R_n(\omega, \cdot)\}$ on $\mathscr{M}_s(\Gamma^\alpha)$ will be shown to have a large deviation property with entropy function $I_{\rho,\alpha}^{(3)}$. Hence $I_{\rho,\alpha}^{(3)}$ is called the *level*-3,α *entropy function*. We now prove that for fixed $\tau \in \mathscr{M}_s(\Gamma^\alpha)$ $I_{\rho,\alpha}^{(3)}(\tau)$ equals the infimum of $I_\rho^{(3)}(P)$ over the set $\mathscr{A}_\alpha = \{P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$: $\pi_\alpha P = \tau\}$. We also show that $I_\rho^{(3)}(P)$ attains its infimum over $\mathscr{A}_\alpha$ at an $(\alpha - 1)$-dependent Markov chain. The latter measures are defined for $\alpha = 2$ in Example A.7.3(b) and for $\alpha \geq 3$ in Appendix A.10.

Lemma IX.3.2. *Let $\alpha \geq 2$ be an integer.*

(a) *A probability measure τ on $\mathscr{B}(\Gamma^\alpha)$ belongs to $\mathscr{M}_s(\Gamma^\alpha)$ if and only if*

$$\sum_{j=1}^r \tau_{i_1 \cdots i_{\alpha-1} j} = \sum_{k=1}^r \tau_{k i_1 \cdots i_{\alpha-1}} \qquad \text{for each } i_1, \ldots, i_{\alpha-1}. \tag{9.7}$$

(b) *Let τ be a measure in $\mathscr{M}_s(\Gamma^\alpha)$. Define $\mathscr{M}_\tau$ to be the subset of $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ consisting of $(\alpha - 1)$-dependent Markov chains which satisfy $\pi_\alpha P = \tau$. Then $\mathscr{M}_\tau$ is nonempty. If $(v_\tau)_{i_1 \cdots i_{\alpha-1}} > 0$ for each $i_1, \ldots, i_{\alpha-1}$, then $\mathscr{M}_\tau$ consists of a unique measure.*

Proof. (a) If τ belongs to $\mathscr{M}_s(\Gamma^\alpha)$, then τ equals $\pi_\alpha P$ for some $P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$, and for each $i_1, \ldots, i_{\alpha-1}$,

$$\sum_{j=1}^r \tau_{i_1 \cdots i_{\alpha-1} j} = P\{\omega \in \Gamma^{\mathbb{Z}} : \omega_1 = x_{i_1}, \ldots, \omega_{\alpha-1} = x_{i_{\alpha-1}}\}$$
$$= \sum_{k=1}^r P\{\omega \in \Gamma^{\mathbb{Z}} : \omega_0 = x_k, \omega_1 = x_{i_1}, \ldots, \omega_{\alpha-1} = x_{i_{\alpha-1}}\}.$$

The last sum equals $\sum_{k=1}^r \tau_{k i_1 \cdots i_{\alpha-1}}$, and so (9.7) holds. Now suppose that (9.7) holds. We write v for v_τ. For each $i_1, \ldots, i_{\alpha-1} \in \{1, \ldots, r\}$, define

$$\gamma_{i_1 \cdots i_\alpha} = \frac{\tau_{i_1 \cdots i_\alpha}}{v_{i_1 \cdots i_{\alpha-1}}}, \qquad i_\alpha = 1, \ldots, r, \qquad \text{if } v_{i_1 \cdots i_{\alpha-1}} > 0.$$

If $v_{i_1 \cdots i_{\alpha-1}} = 0$, then define $\{\gamma_{i_1 \cdots i_\alpha}; i_\alpha = 1, \ldots, r\}$ to be any non-negative real numbers which sum to 1. We have $v_{i_1 \cdots i_{\alpha-1}} \geq 0, \sum_{i_1, \ldots, i_{\alpha-1}=1}^r v_{i_1 \cdots i_{\alpha-1}} = 1$, and by (9.7)

$$\sum_{j=1}^{r} v_{i_1\cdots i_{\alpha-2}j} = \sum_{j,l=1}^{r} \tau_{i_1\cdots i_{\alpha-2}jl} = \sum_{j,k=1}^{r} \tau_{ki_1\cdots i_{\alpha-2}j} = \sum_{k=1}^{r} v_{ki_1\cdots i_{\alpha-2}}.$$

Also $\gamma_{i_1\cdots i_\alpha} \geq 0$, $\sum_{i_\alpha=1}^{r} \gamma_{i_1\cdots i_\alpha} = 1$, $v_{i_1\cdots i_{\alpha-1}}\gamma_{i_1\cdots i_\alpha} = \tau_{i_1\cdots i_\alpha}$ and by (9.7),

$$\sum_{i_1=1}^{r} v_{i_1\cdots i_{\alpha-1}}\gamma_{i_1\cdots i_\alpha} = \sum_{i_1=1}^{r} \tau_{i_1\cdots i_\alpha} = v_{i_2\cdots i_\alpha}.$$

Hence assumptions (A.5) and (A.6) in Appendix A.10 are satisfied. By Theorem A.10.1, there exists a unique $(\alpha - 1)$-dependent Markov chain P in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ with transition array $\gamma = \{\gamma_{i_1\cdots i_\alpha}\}$ and invariant measure $v = \sum_{i_1,\ldots,i_{\alpha-1}=1}^{r} v_{i_1\cdots i_{\alpha-1}}\,\delta_{\{x_{i_1},\ldots,x_{i_{\alpha-1}}\}}$. The marginal $\pi_\alpha P$ equals τ since

$$\begin{aligned}\pi_\alpha P\{x_{i_1},\ldots,x_{i_\alpha}\} &= P\{X_1 = x_{i_1},\ldots,X_\alpha = x_{i_\alpha}\}\\ &= v_{i_1\cdots i_{\alpha-1}}\gamma_{i_1\cdots i_\alpha} = \tau_{i_1\cdots i_\alpha}.\end{aligned}$$

(b) Part (a) shows that $\mathscr{M}_\tau$ is nonempty. Suppose that $v_{i_1\cdots i_{\alpha-1}}$ is positive for each $i_1,\ldots,i_{\alpha-1}$. If P is an $(\alpha - 1)$-dependent Markov chain in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ satisfying $\pi_\alpha P = \tau$, then P must have invariant measure v and transition array $\{\tau_{i_1\cdots i_\alpha}/v_{i_1\cdots i_{\alpha-1}}\}$. We conclude that P is unique. □

Theorem IX.3.3. *Let τ be a measure in $\mathscr{M}_s(\Gamma^\alpha)$, $\alpha \in \{2, 3, \ldots\}$. Define $\mathscr{M}_\tau$ to be the subset of $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ consisting of $(\alpha - 1)$-dependent Markov chains which satisfy $\pi_\alpha P = \tau$. Then $I_\rho^{(3)}(P)$ attains its infimum over the set $\{P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$: $\pi_\alpha P = \tau\}$ at all $P \in \mathscr{M}_\tau$ and only at such P. For any $P \in \mathscr{M}_\tau$*

$$\inf\{I_\rho^{(3)}(P) : P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \pi_\alpha P = \tau\} = I_\rho^{(3)}(P) = I_{\rho,\alpha}^{(3)}(\tau).$$

Proof. Let $\mathscr{A}_\alpha$ be the set $\{P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : \pi_\alpha P = \tau\}$ and P any measure in $\mathscr{A}_\alpha$. If $(v_\tau)_{i_1\cdots i_{\alpha-1}} = p(x_{i_1},\ldots,x_{i_{\alpha-1}}) > 0$, then $p(x_{i_\alpha}|x_{i_1},\ldots,x_{i_{\alpha-1}})$ equals $\tau_{i_1\cdots i_\alpha}/(v_\tau)_{i_1\cdots i_{\alpha-1}}$. Hence

$$H_{\rho,\alpha}(P) = \sum \tau_{i_1\cdots i_\alpha} \log \frac{\tau_{i_1\cdots i_\alpha}}{(v_\tau)_{i_1\cdots i_{\alpha-1}}\rho_{i_\alpha}} = I_{\rho,\alpha}^{(3)}(\tau).$$

By Lemma IX.2.1(c) and Theorem IX.2.3, for any $k \geq \alpha + 1$

$$0 \leq H_{\rho,\alpha}(P) = I_{\rho,\alpha}^{(3)}(\tau) \leq \cdots \leq H_{\rho,k}(P) \leq H_{\rho,k+1}(P) \uparrow I_\rho^{(3)}(P). \tag{9.8}$$

If P is an $(\alpha - 1)$-dependent Markov chain in $\mathscr{M}_\tau$, then for any $k \geq \alpha$ $H_{\rho,k}(P) = I_{\rho,\alpha}^{(3)}(\tau) = I_\rho^{(3)}(P)$. We see that $I_\rho^{(3)}(P)$ attains its infimum over $\mathscr{A}_\alpha$ at $P \in \mathscr{M}_\tau$ and that the infimum equals $I_{\rho,\alpha}^{(3)}(\tau)$. The proof is done once we show that the measures P in $\mathscr{M}_\tau$ are the unique measures in $\mathscr{A}_\alpha$ for which $I_\rho^{(3)}(P)$ equals $I_{\rho,\alpha}^{(3)}(\tau)$.

Suppose that $I_\rho^{(3)}(P)$ equals $I_{\rho,\alpha}^{(3)}(\tau)$ for some P in $\mathscr{A}_\alpha$. We prove by induction that for all $k \geq \alpha$

$$p(x_{i_1},\ldots,x_{i_k}) = (v_\tau)_{i_1\cdots i_{\alpha-1}}\gamma_{i_1\cdots i_\alpha}\cdots\gamma_{i_{k-\alpha+1}\cdots i_k}, \tag{9.9}$$

where the transition array γ is constructed as in Lemma IX.3.2(a). This will imply that P is in $\mathscr{M}_\tau$. Formula (9.9) holds for $k = \alpha$ since $\pi_\alpha P$ equals τ. Since $I_\rho^{(3)}(P) = I_{\rho,\alpha}^{(3)}(\tau)$, (9.8) shows that $H_{\rho,k}(P)$ equals $H_{\rho,\alpha}(P)$ for all $k \geq \alpha + 1$. Assume that (9.9) has been shown for $k = \alpha, \alpha + 1, \ldots, c - 1$, some $c \geq \alpha + 1$. With respect to P, X_1 and X_c are conditionally independent given $X_2, \ldots, X_{c-1}$ [Lemma IX.2.1(e)]. If $p(x_{i_1}, \ldots, x_{i_{c-1}}) > 0$, then by the induction hypothesis and strict stationarity,

$$p(x_{i_1}, \ldots, x_{i_c}) = p(x_{i_1}, \ldots, x_{i_{c-1}})p(x_{i_c}|x_{i_2}, \ldots, x_{i_{c-1}})$$
$$= (v_\tau)_{i_1 \cdots i_{\alpha-1}} \gamma_{i_1 \cdots i_\alpha} \cdots \gamma_{i_{c-\alpha} \cdots i_{c-1}} \gamma_{i_{c-\alpha+1} \cdots i_c}.$$

If $p(x_{i_1}, \ldots, x_{i_{c-1}}) = 0$, then

$$p(x_{i_1}, \ldots, x_{i_c}) = 0 = (v_\tau)_{i_1 \cdots i_{\alpha-1}} \gamma_{i_1 \cdots i_\alpha} \cdots \gamma_{i_{c-\alpha+1} \cdots i_c}.$$

Thus (9.9) holds for $k = c$. The proof of the theorem is complete. □

The contraction principles just proved will now be used in order to deduce the level-3 large deviation bounds.

IX.4. Proof of the Level-3 Large Deviation Bounds

In this section we prove that

$$\limsup_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{K\} \leq -I_\rho^{(3)}(K) \qquad \text{for each closed set } K \text{ in } \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \tag{9.10}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{G\} \geq -I_\rho^{(3)}(G) \qquad \text{for each open set } G \text{ in } \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \tag{9.11}$$

where $Q_n^{(3)}$ is the P_ρ-distribution of $R_n(\omega, \cdot)$ on $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. With these bounds, we will complete the proof of the level-3 large deviation property since $I_\rho^{(3)}$ has already been shown to be lower semicontinuous and to have compact level sets. Our strategy is first to show that for each $\alpha \geq 2$ the P_ρ-distributions of the α-dimensional marginals $\{\pi_\alpha R_n(\omega, \cdot)\}$ have a large deviation property with entropy function $I_{\rho,\alpha}^{(3)}$ defined in (9.6). The bounds (9.10) and (9.11) will follow by an approximation argument. We prove the large deviation property for $\{\pi_\alpha R_n(\omega, \cdot)\}$ by applying the large deviation theorem for random vectors, Theorem II.6.1. In order to calculate the corresponding free energy functions, we need some facts about non-negative matrices.

Let $B = \{B_{ij}\}$ be a real, square matrix. We say that B is *non-negative*, and write $B \geq 0$, if each B_{ij} is non-negative. We say that B is *positive*, and write $B > 0$, if each B_{ij} is positive. If B is non-negative, then B is said to be *primitive* if there exists a positive integer k such that B^k is positive. Clearly, a positive matrix is primitive. If B is non-negative, then B is said to be *stochastic* if $\sum_j B_{ij} = 1$ for each i. The next lemma is due to Perron and Frobenius.

Lemma IX.4.1. *Let $B = \{B_{ij}\}$ be a non-negative primitive matrix. Then there exists an eigenvalue $\lambda(B)$ of B with the following properties.*

(a) *$\lambda(B)$ is real and positive and $\lambda(B)$ exceeds in absolute value any other eigenvalue of B.*

(b) *$\lambda(B)$ is a simple root of the characteristic equation of B.*

(c) *With $\lambda(B)$ may be associated a positive left eigenvector u and a positive right eigenvector w. These eigenvectors are unique up to constant multiples.*

(d) $\lim_{n\to\infty} n^{-1}\log B_{ij}^n = \log\lambda(B)$ *for each i and j.**

(e) *If B is stochastic, then $\lambda(B)$ equals* 1.

(f) *If the entries of B are $\mathscr{C}^1$ functions of a parameter $t\in\mathbb{R}^d$, then $\lambda(B)$ is a $\mathscr{C}^1$ function of $t\in\mathbb{R}^d$; in particular, $\lambda(B)$ is a differentiable function of $t\in\mathbb{R}^d$.*

Proof. (a)–(c) See, e.g., Seneta (1981, Theorem 1.1).

(d) Let u and w be positive left and right eigenvectors associated with $\lambda(B)$ and normalized so that $\langle u, w\rangle = 1$ [part (c)]. Define $\gamma_{ij} = B_{ij}w_j/(\lambda(B)w_i)$. The matrix $\gamma = \{\gamma_{ij}\}$ is a stochastic matrix which is primitive since B is primitive. Since the vector $v_i = u_i w_i$ satisfies $\sum_i v_i = 1$ and $\sum_i v_i\gamma_{ij} = v_j$ for each j, it follows that $\gamma_{ij}^n \to v_j = u_j w_j$ for each i and j [Lemma A.9.5]. This limit and the fact that

$$\gamma_{ij}^n = B_{ij}^n w_j/(\lambda(B)^n w_i), \qquad n = 1, 2, \ldots,$$

imply that $n^{-1}\log B_{ij}^n \to \log\lambda(B)$ as $n\to\infty$.

(e) Since $\sum_j B_{ij}1 = 1$, $\lambda(B)$ cannot be less than 1. If w is a positive right eigenvector corresponding to $\lambda(B)$, then pick an index i such that $w_i = \max_j w_j$. We have

$$\lambda(B) = \sum_j B_{ij}w_j/w_i \le \sum_j B_{ij}\max_j w_j/\max_j w_j = 1.$$

(f) The eigenvalue $\lambda = \lambda(B)$ is a simple root of the characteristic equation $\det(\lambda\delta_{ij} - B_{ij}(t)) = 0$, in which the entries $B_{ij}(t)$ are $\mathscr{C}^1$ functions of $t\in\mathbb{R}^d$. The implicit function theorem completes the proof. □

If $\alpha \ge 2$ is an integer, then $\mathscr{M}_s(\Gamma^\alpha)$ denotes the set of probability measures τ on $\mathscr{B}(\Gamma^\alpha)$ which are of the form $\tau = \pi_\alpha P$ for some P in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. $\mathscr{M}_s(\Gamma^\alpha)$ with the topology of weak convergence is homeomorphic to a compact convex subset $\mathscr{M}_{s,\alpha}$ of $\mathbb{R}^{r^\alpha}$. $\mathscr{M}_{s,\alpha}$ consists of all vectors $\tau = \{\tau_{i_1\cdots i_\alpha}; i_1, \ldots, i_\alpha = 1, \ldots, r\}$ which satisfy $\tau_{i_1\cdots i_\alpha} \ge 0$ for each $i_1, \ldots, i_\alpha$, $\sum_{i_1,\ldots,i_\alpha=1}^r \tau_{i_1\cdots i_\alpha} = 1$, and

$$(9.12) \qquad \sum_{j=1}^r \tau_{i_1\cdots i_{\alpha-1}j} = \sum_{k=1}^r \tau_{ki_1\cdots i_{\alpha-1}} \qquad \text{for each } i_1, \ldots, i_{\alpha-1}.$$

For $\tau = \{\tau_{i_1,\ldots,i_\alpha}\}$ a point in $\mathbb{R}^{r^\alpha}$, we define the function

$$(9.13) \qquad \bar{I}_{\rho,\alpha}^{(3)}(\tau) = \begin{cases} I_{\rho,\alpha}^{(3)}(\tau) & \text{for } \tau\in\mathscr{M}_{s,\alpha}, \\ \infty & \text{for } \tau\notin\mathscr{M}_{s,\alpha}, \end{cases}$$

where $I_{\rho,\alpha}^{(3)}$ is defined in (9.6). $\bar{I}_{\rho,\alpha}^{(3)}$ is continuous relative to $\mathscr{M}_{s,\alpha}$.

* B_{ij}^n denotes the ij-entry of the product B^n.

Large deviation property of $\{\pi_1 R_n(\omega, \cdot)\}$. Denote by $I_{\rho,1}^{(3)}$ the relative entropy function $I_\rho^{(2)}$. The one-dimensional marginal $\pi_1 R_n(\omega, \cdot)$ equals the empirical measure $L_n(\omega, \cdot)$. The following theorem was proved in Section VIII.2.

Theorem IX.4.2. *The* P_ρ*-distributions of* $\{\pi_1 R_n(\omega, \cdot); n = 1, 2, \ldots,\}$ on $\mathcal{M}(\Gamma)$ *have a large deviation property with* $a_n = n$ *and entropy function* $I_{\rho,1}^{(3)} = I_\rho^{(2)}$.

This was proved by showing a large deviation property for the random vectors $L_n(\omega) = (L_n(\omega, \{x_1\}), \ldots, L_n(\omega, \{x_r\}))$. We now prove a generalization for the α-dimensional marginals $\{\pi_\alpha R_n(\omega, \cdot)\}$, $\alpha \geq 2$; $\pi_\alpha R_n(\omega, \cdot)$ takes values in the space $\mathcal{M}_s(\Gamma^\alpha)$.

Theorem IX.4.3. *For* $\alpha \geq 2$, *the* P_ρ*-distributions of* $\{\pi_\alpha R_n(\omega, \cdot); n = 1, 2, \ldots\}$ *on* $\mathcal{M}_s(\Gamma^\alpha)$ *have a large deviation property with* $a_n = n$ *and entropy function* $I_{\rho,\alpha}^{(3)}$.

Let $M_{n,\alpha}(\omega)$ denote the vector in $\mathbb{R}^{r^\alpha}$ with

$$(M_{n,\alpha}(\omega))_{i_1 \cdots i_\alpha} = \pi_\alpha R_n(\omega, \{x_{i_1}, \ldots, x_{i_\alpha}\}), \qquad i_1, \ldots, i_\alpha = 1, \ldots, r.$$

$M_{n,\alpha}(\omega)$ takes values in the compact convex subset $\mathcal{M}_{s,\alpha}$ of $\mathbb{R}^{r^\alpha}$. Exactly as in Section VIII.2, it suffices to prove that the distributions of $\{M_{n,\alpha}\}$ on $\mathbb{R}^{r^\alpha}$ have a large deviation property with entropy function $\bar{I}_{\rho,\alpha}^{(3)}(\tau)$. We apply Theorem II.6.1. It is convenient to divide the proof into three steps.

Step 1: *Evaluation of the free energy function.* Let us first consider $\alpha = 2$. We write

$$(M_{n,2}(\omega))_{ij} = \frac{1}{n} \sum_{\beta=1}^{n} \delta_{Y_\beta^{(n)}(\omega)}\{x_i, x_j\},$$

where $Y_\beta^{(n)}(\omega)$ equals the ordered pair $(X_\beta(\omega), X_{\beta+1}(\omega))$ for $\beta \in \{1, \ldots, n-1\}$ and equals the cyclic pair $(X_n(\omega), X_1(\omega))$ for $\beta = n$. If $t = \{t_{ij}; i, j = 1, \ldots, r\}$ is a point in $\mathbb{R}^{r^2}$, then define the function $f_t(x_i, x_j) = t_{ij}$ for $x_i, x_j \in \Gamma$. Thus $f_t(Y_\beta^{(n)}(\omega)) = t_{ij}$ if $Y_\beta^{(n)}(\omega) = (x_i, x_j)$ and

$$\langle t, M_{n,2}(\omega) \rangle = \frac{1}{n} \sum_{i,j=1}^{r} t_{ij} \sum_{\beta=1}^{n} \delta_{Y_\beta^{(n)}(\omega)}\{x_i, x_j\} = \frac{1}{n} \sum_{\beta=1}^{n} f_t(Y_\beta^{(n)}(\omega)).$$

In the notation of Theorem II.6.1, W_n equals $nM_{n,2}$ and a_n equals n. The free energy function $c_2(t)$ of the sequence $\{nM_{n,2}(\omega); n = 1, 2, \ldots\}$ equals $\lim_{n\to\infty} c_{n,2}(t)$, where $c_{n,2}(t)$ is given by

$$\begin{aligned}\frac{1}{n}\log E_\rho\{\exp(n\langle t, M_{n,2}\rangle)\} &= \frac{1}{n}\log E_\rho\left\{\exp \sum_{\beta=1}^{n} f_t(Y_\beta^{(n)})\right\} \\ &= \frac{1}{n}\log \sum_{i_1, \ldots, i_n=1}^{r} \exp(t_{i_1 i_2} + \cdots + t_{i_{n-1} i_n} \\ &\qquad + t_{i_n i_1}) \cdot \rho_{i_1} \cdots \rho_{i_n}.\end{aligned}$$

Define $B_2(t)$ to be the positive matrix $\{e^{t_{ij}}\rho_j; i,j = 1, \ldots, r\}$. Then

$$c_{n,2}(t) = \frac{1}{n}\log \sum_{i_1=1}^{r} B_2(t)^n_{i_1 i_1}.$$

By Lemma IX.4.1(d), $c_2(t) = \lim_{n\to\infty} c_{n,2}(t) = \log \lambda(B_2(t))$.

In order to evaluate the free energy function $c_\alpha(t)$ of the sequence $\{nM_{n,\alpha}\}$ for $\alpha \geq 3$, we need some notation. If $n \geq \alpha$ and if $i_1, i_2, \ldots, i_n \in \{1, \ldots, r\}$, then define the multi-index

$$i(n,\alpha,j) = \begin{cases} (i_{j+1}, i_{j+2}, \ldots, i_{j+\alpha}) & \text{for } 0 \leq j \leq n-\alpha, \\ (i_{j+1}, \ldots, i_n, i_1, \ldots, i_{j+\alpha-n}) & \text{for } n-\alpha+1 \leq j \leq n-1. \end{cases}$$

Given $t = \{t_{i_1 \ldots i_\alpha}; i_1, \ldots, i_\alpha = 1, \ldots, r\}$ a point in $\mathbb{R}^{r^\alpha}$, let $B_\alpha(t)$ be the $r^{\alpha-1} \times r^{\alpha-1}$ matrix

$$(9.14)\qquad B_\alpha(t)_{i_1\ldots i_{\alpha-1}, j_1 \ldots j_{\alpha-1}} = \begin{cases} \exp(t_{i_1 i_2 \ldots i_{\alpha-1} j_{\alpha-1}})\rho_{j_{\alpha-1}} & \text{for } 1 \leq i_1, i_2 = j_1, i_3 = j_2, \ldots, \\ & \quad i_{\alpha-1} = j_{\alpha-2}, j_{\alpha-1} \leq r, \\ 0 & \text{otherwise.} \end{cases}$$

Then $c_\alpha(t)$ equals $\lim_{n\to\infty} c_{n,\alpha}(t)$, where for $n \geq \alpha$ $c_{n,\alpha}(t)$ is given by

$$(9.15)\qquad \begin{aligned} \frac{1}{n}\log E_\rho\{\exp(n\langle t, M_{n,\alpha}\rangle)\} &= \frac{1}{n}\log \sum_{i_1,\ldots,i_n=1}^{r} \exp\left(\sum_{j=0}^{n-1} t_{i(n,\alpha,j)}\right)\rho_{i_1}\cdots\rho_{i_n} \\ &= \frac{1}{n}\log \sum_{i_1,\ldots,i_{\alpha-1}=1}^{r} B_\alpha(t)^n_{i_1\ldots i_{\alpha-1}, i_1 \ldots i_{\alpha-1}}. \end{aligned}$$

$B = B_\alpha(t)$ is primitive since

$$B^{\alpha-1}_{i_1\ldots i_{\alpha-1}, j_1\ldots j_{\alpha-1}} = B_{i_1\ldots i_{\alpha-1}, i_2\ldots i_{\alpha-1} j_1} B_{i_2\ldots i_{\alpha-1} j_1, i_3 \ldots i_{\alpha-1} j_1 j_2} \cdots B_{i_{\alpha-1} j_1 \ldots j_{\alpha-2}, j_1 \ldots j_{\alpha-2} j_{\alpha-1}} > 0.$$

By Lemma IX.4.1(d), $c_\alpha(t) = \lim_{n\to\infty} c_{n,\alpha}(t) = \log\lambda(B_\alpha(t))$.

Step 2: *Large deviation property*. The free energy function of the sequence $\{nM_{n,\alpha}; n = 1, 2, \ldots\}$ equals $\log\lambda(B_\alpha(t))$. Since the entries of $B_\alpha(t)$ are $\mathscr{C}^1$ functions of $t \in \mathbb{R}^{r^\alpha}$, $\lambda(B_\alpha(t))$ is a differentiable function of $t \in \mathbb{R}^{r^\alpha}$ [Lemma IX.4.1(f)]; $\log\lambda(B_\alpha(t))$ has the same property. By Theorem II.6.1, the P_ρ-distributions of $\{M_{n,\alpha}\}$ on $\mathbb{R}^{r^\alpha}$ have a large deviation property with entropy function

$$(9.16)\qquad I_{\rho,\alpha}(\tau) = \sup_{t\in\mathbb{R}^{r^\alpha}} \{\langle t, \tau\rangle - \log\lambda(B_\alpha(t))\}, \qquad \tau \in \mathbb{R}^{r^\alpha}.$$

If $Q_{n,\alpha}$ denotes the P_ρ-distribution of $M_{n,\alpha}$, then

$$(9.17)\qquad \limsup_{n\to\infty} \frac{1}{n}\log Q_{n,\alpha}\{K\} \leq -I_{\rho,\alpha}(K) \qquad \text{for each closed set } K \text{ in } \mathbb{R}^{r^\alpha},$$

(9.18) $$\liminf_{n\to\infty}\frac{1}{n}\log Q_{n,\alpha}\{G\} \geq -I_{\rho,\alpha}(G) \qquad \text{for each open set } G \text{ in } \mathbb{R}^{r^\alpha}.$$

Step 3: *Evaluation of the entropy function.* We show that for $\alpha \geq 2$ and $\tau \in \mathbb{R}^{r^\alpha}$, $I_{\rho,\alpha}(\tau)$ defined in (9.16) equals $\bar{I}^{(3)}_{\rho,\alpha}(\tau)$ defined in (9.13). In other words, we prove that

(9.19) $$\bar{I}^{(3)}_{\rho,\alpha}(\tau) = \sup_{t\in\mathbb{R}^{r^\alpha}}\{\langle t,\tau\rangle - \log\lambda(B_\alpha(t))\} \qquad \text{for } \tau\in\mathbb{R}^{r^\alpha}.$$

If G is any open subset of $\mathbb{R}^{r^\alpha}$ which is disjoint from $\mathcal{M}_{s,\alpha}$, then $Q_{n,\alpha}\{G\}$ equals 0, and by the lower bound (9.18) $I_{\rho,\alpha}(G)$ equals ∞. Thus for $\tau\notin\mathcal{M}_{s,\alpha}$, $I_{\rho,\alpha}(\tau) = \infty = \bar{I}^{(3)}_{\rho,\alpha}(\tau)$.

Since $I_{\rho,\alpha}(\tau)$ is defined as a Legendre–Fenchel transform, $I_{\rho,\alpha}(\tau)$ is a closed convex function on $\mathbb{R}^{r^\alpha}$. Suppose we show that $I_{\rho,\alpha}(\tau)$ equals $\bar{I}^{(3)}_{\rho,\alpha}(\tau)$ for all τ in the relative interior of $\mathcal{M}_{s,\alpha}$.* Since $\bar{I}^{(3)}_{\rho,\alpha}$ is continuous relative to $\mathcal{M}_{s,\alpha}$, the continuity property in Theorem VI.3.2 will imply that $I_{\rho,\alpha}(\tau)$ equals $\bar{I}^{(3)}_{\rho,\alpha}(\tau)$ for all τ in $\mathcal{M}_{s,\alpha}$. This will complete the proof of the large deviation property of $\{\pi_\alpha R_n(\omega,\cdot)\}$ with entropy function $I^{(3)}_{\rho,\alpha}$.

In order to treat all values of α by a single proof, it is useful to introduce multi-indices $i=(i_1,\ldots,i_{\alpha-1})$ and $j=(j_1,\ldots,j_{\alpha-1})$, where $i_1,\ldots,i_{\alpha-1}$, $j_1,\ldots,j_{\alpha-1}\in\{1,\ldots,r\}$. For $\alpha>2$, we write $i\sim j$ if all the equalities $i_2=j_1$, $i_3=j_2,\ldots,i_{\alpha-1}=j_{\alpha-2}$ hold. Otherwise we write $i\not\sim j$. For $\alpha=2$, we write $i\sim j$ for any i and j. For $\tau\in\mathcal{M}_{s,\alpha}$, define τ_{ij} to be $\tau_{i_1\cdots i_{\alpha-1}j_{\alpha-1}}$ if $i\sim j$ and to be 0 if $i\not\sim j$. Set $(v_\tau)_i = \sum_j \tau_{ij}$. For $t\in\mathbb{R}^{r^\alpha}$, define t_{ij} like τ_{ij}. In this notation

$$\bar{I}^{(3)}_{\rho,\alpha}(\tau) = \sum \tau_{ij}\log\frac{\tau_{ij}}{(v_\tau)_i\rho_{j_{\alpha-1}}} \qquad \text{and} \qquad B_\alpha(t)_{ij} = \begin{cases} e^{t_{ij}}\rho_{j_{\alpha-1}} & \text{if } i\sim j,\\ 0 & \text{if } i\not\sim j.\end{cases}$$

The sum defining $\bar{I}^{(3)}_{\rho,\alpha}(\tau)$ runs over all i and j for which $(v_\tau)_i>0$ and $i\sim j$. Now let τ^0 be any point in the relative interior of $\mathcal{M}_{s,\alpha}$. Then τ^0_{ij} is positive whenever $i\sim j$, and so each $(v_{\tau^0})_i$ is positive. If we define

$$t^0_{ij} = \log\frac{\tau^0_{ij}}{(v_{\tau^0})_i\rho_{j_{\alpha-1}}} \qquad \text{for all } i\sim j,$$

then $B_\alpha(t^0)_{ij} = \tau^0_{ij}/(v^0_\tau)_i$ if $i\sim j$ and $B_\alpha(t^0)_{ij}=0$ if $i\not\sim j$. $B_\alpha(t^0)$ is a stochastic matrix. Hence $\log\lambda(B_\alpha(t^0))$ equals 0 [Lemma IX.4.1(e)] and

$$I_{\rho,\alpha}(\tau^0) \geq \langle t^0,\tau^0\rangle - \log\lambda(B_\alpha(t^0)) = \sum_{i\sim j}\tau^0_{ij}\log\frac{\tau^0_{ij}}{(v^0_\tau)_i\rho_{j_{\alpha-1}}} = \bar{I}^{(3)}_{\rho,\alpha}(\tau^0).$$

In order to prove that $\bar{I}^{(3)}_{\rho,\alpha}(\tau^0) \geq I_{\rho,\alpha}(\tau^0) = \sup_{t\in\mathbb{R}^{r^\alpha}}\{\langle t,\tau^0\rangle - \log\lambda(B_\alpha(t))\}$, it suffices to prove that $\log\lambda(B_\alpha(t)) \geq \langle t,\tau^0\rangle - \bar{I}^{(3)}_{\rho,\alpha}(\tau^0)$ for all $t\in\mathbb{R}^{r^\alpha}$. We will in fact show that for all $t\in\mathbb{R}^{r^\alpha}$

(9.20) $$\log\lambda(B_\alpha(t)) = \sup_{\tau\in \operatorname{ri}\mathcal{M}_{s,\alpha}}\{\langle t,\tau\rangle - \bar{I}^{(3)}_{\rho,\alpha}(\tau)\}.$$

*The relative interior of $\mathcal{M}_{s,\alpha}$ consists of all positive vectors τ in $\mathcal{M}_{s,\alpha}$ (each $\tau_{i_1\cdots i_\alpha}>0$) [Rockafellar (1970, page 48)].

If we insert the definition of $\bar{I}^{(3)}_{\rho,\alpha}(\tau)$, then the right-hand side of (9.20) becomes

$$\sup_{\tau \in \operatorname{ri}\mathcal{M}_{s,\alpha}} \sum_{i \sim j} \tau_{ij} \log \frac{B_\alpha(t)_{ij}(v_\tau)_i}{\tau_{ij}}.$$

The matrix $B_\alpha(t)$ is primitive and $B_\alpha(t)_{ij} > 0$ if and only if $i \sim j$. For any τ belonging to the relative interior of $\mathcal{M}_{s,\alpha}$, we have $\tau_{ij} \geq 0$, $\tau_{ij} > 0$ if and only if $B_\alpha(t)_{ij} > 0$, $\sum_{i,j} \tau_{ij} = 1$, and $\sum_j \tau_{ij} = \sum_k \tau_{ki}$ for each i. Hence (9.20) is a consequence of the next theorem.

Theorem IX.4.4. *Let $B = \{B_{ij}\}$ be a non-negative, primitive, $m \times m$ matrix (some $m \geq 2$). Let $\mathcal{M}_B$ be the set of all $\tau = \{\tau_{ij}; i,j = 1, \ldots, m\}$ which satisfy $\tau_{ij} \geq 0$, $\tau_{ij} > 0$ if and only if $B_{ij} > 0$, $\sum_{i,j=1}^m \tau_{ij} = 1$, and $\sum_{j=1}^m \tau_{ij} = \sum_{k=1}^m \tau_{ki}$ for each i. Define $(v_\tau)_i = \sum_{j=1}^m \tau_{ij}$. Then*

$$\log \lambda(B) = \sup_{\tau \in \mathcal{M}_B} \sum\nolimits_B \tau_{ij} \log \frac{B_{ij}(v_\tau)_i}{\tau_{ij}}, \tag{9.21}$$

where $\sum_B$ denotes the sum over all i and j for which $B_{ij} > 0$.

Proof. Let u and w be positive left and right eigenvectors associated with $\lambda(B)$ and normalized so that $\langle u, w \rangle = 1$. Define $\tau^0_{ij} = u_i B_{ij} w_j / \lambda(B)$ and $v^0_i = \sum_{j=1}^m \tau^0_{ij}$. $\tau^0 = \{\tau^0_{ij}\}$ belongs to $\mathcal{M}_B$. We prove that the supremum in (9.21) is attained at the unique point τ^0. Since v^0_i equals $u_i w_i$,

$$\sum\nolimits_B \tau^0_{ij} \log \frac{B_{ij} v^0_i}{\tau^0_{ij}} = \sum\nolimits_B \tau^0_{ij} \log \frac{w_i \lambda(B)}{w_j} = \log \lambda(B). \tag{9.22}$$

Let $\tau \neq \tau^0$ be any other point in $\mathcal{M}_B$. Each $(v_\tau)_i$ is positive. Indeed, if $(v_\tau)_i = 0$, then $\tau_{ij} = 0 = \tau_{ki}$ for each j and k, and so $B_{ij} = 0 = B_{ki}$ for each j and k. The latter cannot hold since B is primitive. Set $v = v_\tau$, $\gamma_{ij} = \tau_{ij}/v_i$, and $\gamma^0_{ij} = \tau^0_{ij}/v^0_i$. We have

$$\sum\nolimits_B \tau_{ij} \log \frac{B_{ij} v_i}{\tau_{ij}} = \sum\nolimits_B \tau_{ij} \log \frac{B_{ij} v^0_i}{\tau^0_{ij}} - \sum\nolimits_B \tau_{ij} \log\left(\frac{\tau_{ij}}{v_i} \frac{v^0_i}{\tau^0_{ij}}\right).$$

By the same calculation as in (9.22), the first sum equals $\log \lambda(B)$. Since $x \log x \geq x - 1$ with equality iff $x = 1$,

$$-\sum\nolimits_B \tau_{ij} \log\left(\frac{\tau_{ij}}{v_i} \frac{v^0_i}{\tau^0_{ij}}\right) = -\sum\nolimits_B \frac{v_i \tau^0_{ij}}{v^0_i} \frac{\tau_{ij} v^0_i}{v_i \tau^0_{ij}} \log\left(\frac{\tau_{ij} v^0_i}{v_i \tau^0_{ij}}\right) \leq -\sum\nolimits_B \left(\tau_{ij} - \frac{v_i \tau^0_{ij}}{v^0_i}\right) = 0 \tag{9.23}$$

and equality holds iff $\gamma_{ij} = \gamma^0_{ij}$ for each i and j. Since $\{\gamma_{ij}\}$ and $\{\gamma^0_{ij}\}$ are primitive stochastic matrices, $\{\gamma_{ij}\} = \{\gamma^0_{ij}\}$ implies that v and v^0 are equal [Lemma A.9.5]. Hence equality holds in (9.23) iff $\tau = \tau^0$. We conclude that for $\tau \in \mathcal{M}_B$, $\sum_B \tau_{ij} \log(B_{ij} v_i / \tau_{ij}) \leq \log \lambda(B)$ with equality iff $\tau = \tau^0$. □

With the last theorem, we have completed the proof of Theorem IX.4.3 (large deviation property of $\{\pi_\alpha R_n(\omega, \cdot)\}$, $\alpha \geq 2$). We are ready to prove the large deviation bounds (9.10) and (9.11) for the P_ρ-distributions of $\{R_n(\omega, \cdot)\}$ on $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. These bounds will be derived by means of an approximation argument based on our previous work in this chapter.

If A is any subset of $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$, then obviously $P \in A$ implies $\pi_\alpha P \in \pi_\alpha A$ for any $\alpha \geq 1$. We say that A is *finite dimensional* if for all sufficiently large α, $\pi_\alpha P \in \pi_\alpha A$ implies $P \in A$.

Example IX.4.5. Let $\Sigma_1, \ldots, \Sigma_l$ be cylinder sets, P an element of $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$, and $\varepsilon > 0$. Consider the open set

$$G_0 = \{\bar{P} \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : |\bar{P}\{\Sigma_i\} - P\{\Sigma_i\}| < \varepsilon, i = 1, \ldots, l\}.$$

Since we are dealing with strictly stationary measures, we may assume without loss of generality that each Σ_i has the form

$$\Sigma_i = \{\omega \in \Gamma^{\mathbb{Z}} : (\omega_1, \ldots, \omega_{\alpha_i}) \in F_i\},$$

where α_i is a positive integer and F_i is a subset of Γ^{α_i}. If $\alpha \geq \max_{i=1,\ldots,l} \alpha_i$, then $\pi_\alpha^{-1}(\pi_\alpha \Sigma_i) = \Sigma_i$ and so

$$G_0 = \{\bar{P} \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : |\pi_\alpha \bar{P}\{\pi_\alpha \Sigma_i\} - P\{\Sigma_i\}| < \varepsilon, i = 1, \ldots, l\}.$$

This shows that G_0 is finite dimensional.

For such open sets G_0, we can derive the lower large deviation bound (9.11) using the large deviation property of $\{\pi_\alpha R_n(\omega, \cdot)\}$ with entropy function $I_{\rho,\alpha}^{(3)}$ [Theorems IX.4.2 and IX.4.3] and the contraction principle relating $I_{\rho,\alpha}^{(3)}$ and $I_\rho^{(3)}$ [Theorems IX.3.1 and IX.3.3]. Given such a set G_0, pick α such that $\pi_\alpha P \in \pi_\alpha G_0$ implies $P \in G_0$. Then

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{G_0\} \\
&= \liminf_{n\to\infty} \frac{1}{n} \log P_\rho\{\pi_\alpha R_n \in \pi_\alpha G_0\} && (G_0 \text{ finite dimensional}) \\
&\geq -I_{\rho,\alpha}^{(3)}(\pi_\alpha G_0) && (\pi_\alpha G_0 \text{ open; large deviation property of } \pi_\alpha R_n) \\
&= -\inf\{I_\rho^{(3)}(P) : P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \pi_\alpha P \in \pi_\alpha G_0\} && (\text{contraction principle}) \\
&= -I_\rho^{(3)}(G_0) && (G_0 \text{ finite dimensional}).
\end{aligned}$$

This is (9.11). A similar proof yields the upper bound (9.10) for closed sets of the form $K_0 = \{\bar{P} \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : |\bar{P}\{\Sigma_i\} - P\{\Sigma_i\}| \leq \varepsilon, i = 1, \ldots, l\}$.

We now prove the lower large deviation bound (9.11) for any nonempty open set G in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. For the empty set, (9.11) holds trivially. An open base for the topology on $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ is the class of sets of the form

$$\left\{\bar{P} \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : \left| \int_{\Gamma^{\mathbb{Z}}} f_i \, d\bar{P} - \int_{\Gamma^{\mathbb{Z}}} f_i \, dP \right| < \varepsilon, i = 1, \ldots, k\right\}, \tag{9.24}$$

where $P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$, $f_1, \ldots, f_k \in \mathscr{C}(\Gamma^{\mathbb{Z}})$, and $\varepsilon > 0$. Let $\mathscr{A}$ denote the subset of $\mathscr{C}(\Gamma^{\mathbb{Z}})$ consisting of all finite, real, linear combinations $\sum a_i \chi_{\Sigma_i}$, where $\{\Sigma_i\}$ are cylinder sets. By the Stone–Weierstrass theorem [Theorem A.11.1], $\mathscr{A}$ is dense in $\mathscr{C}(\Gamma^{\mathbb{Z}})$. Hence each set (9.24) contains a set of the form

$$(9.25) \qquad G_0 = \{\bar{P} \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : |\bar{P}\{\Sigma_i\} - P\{\Sigma_i\}| < \varepsilon, i = 1, \ldots, l\},$$

where $\Sigma_1, \ldots, \Sigma_l$ are cylinder sets. Since each set (9.25) is also a set of the form (9.24) (with $k = l$ and $f_i = \chi_{\Sigma_i}$), it follows that the class of sets (9.25) is an open base for the topology on $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. Each set G_0 in (9.25) is open and finite dimensional. If G in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ is nonempty and open and $G = \cup G_0$, then for each set G_0 contained in G

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{G\} \geq \liminf_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{G_0\} \geq -I_\rho^{(3)}\{G_0\}.$$

This follows from the lower large deviation bound for the open, finite-dimensional set G_0. In the last display, we may replace $-I_\rho^{(3)}(G_0)$ by $\sup_{G_0 \subset G}\{-I_\rho^{(3)}(G_0)\}$, and we conclude that

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{G\} \geq \sup_{G_0 \subset G} \{-I_\rho^{(3)}(G_0)\} &= -\inf_{G_0 \subset G} I_\rho^{(3)}(G_0) \\ &= -I_\rho^{(3)}(\cup G_0) \\ &= -I_\rho^{(3)}(G).\end{aligned}$$

This is (9.11).

We end by proving the upper large deviation bound (9.10) for any nonempty closed set K in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$. For the empty set, (9.10) holds trivially. Let $\delta > 0$ be given. We have just seen that any nonempty open set in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ contains a finite dimensional set of the form (9.25). Since $I_\rho^{(3)}$ is lower semicontinuous, it follows that for each $P \in K$ there exist cylinder sets $\Sigma_1, \ldots, \Sigma_l$ (depending on P) and a positive real number ε_P such that if $\bar{P}$ is contained in the set

$$K_P = \{\bar{P} \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}) : |\bar{P}\{\Sigma_i\} - P\{\Sigma_i\}| \leq \varepsilon_P, i = 1, \ldots, l\},$$

then $I_\rho^{(3)}(\bar{P}) > I_\rho^{(3)}(P) - \delta$.* We have $K \subseteq \bigcup_{P \in K} K_P$, and since $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ is compact, K is compact [Lemma A.9.2(c)]. Hence there exist finitely many elements $P_1, \ldots, P_s$ of K such that $K \subseteq \bigcup_{i=1}^s K_{P_i}$. Each set K_{P_i} is closed and finite dimensional. By the upper large deviation bound for K_{P_i},

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{K\} &\leq \limsup_{n\to\infty} \frac{1}{n} \log\left(\sum_{j=1}^s Q_n^{(3)}\{K_{P_i}\}\right) \\ &= \max_{i=1,\ldots,s} \limsup_{n\to\infty} \frac{1}{n} \log Q_n^{(3)}\{K_{P_i}\}\end{aligned}$$

* $I_\rho^{(3)}(P) < \infty$ for $P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}})$ [Theorem IX.2.3].

$$\leq \max_{i=1,\ldots,s} \{-I_\rho^{(3)}(K_{P_i})\}$$

$$\leq \max_{i=1,\ldots,s} \{-I_\rho^{(3)}(P_i) + \delta\}$$

$$\leq -I_\rho^{(3)}(K) + \delta.$$

We obtain (9.10) upon letting δ tend to 0. This completes the proof of the level-3 large deviation property.

IX.5. Notes

1 (page 269). The methods of this chapter can be used to prove large deviation properties for Markov chains with a finite state space. The level-1 property is given as Problem IX.6.6, the level-2 property as Problem IX.6.7, and the level-3 property as Problems IX.6.10–IX.6.15. Donsker and Varadhan (1985) prove a level-3 large deviation property for Gaussian processes. Orey (1985) studies level-3 large deviation problems related to dynamical systems.

2 (page 273) Theorem IX.2.3 showed that $\lim_{\alpha\to\infty} \alpha^{-1} I_{\pi_\alpha P}^{(2)}(\pi_\alpha P)$ exists and equals the quantity $I_\rho^{(3)}(P)$ defined in (9.1). Here is an elementary proof of just the existence of $\lim_{\alpha\to\infty} \alpha^{-1} I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P)$. Using the notation on page 271, we define

$$H_\alpha(P) = \begin{cases} -\sum_{i_1=1}^{r} p(x_{i_1}) \log p(x_{i_1}) & \text{for } \alpha = 1, \\ -\sum p(x_{i_1}, \ldots, x_{i_\alpha}) \log p(x_{i_\alpha} | x_{i_1}, \ldots, x_{i_{\alpha-1}}) & \text{for } \alpha \geq 2. \end{cases}$$

For $\alpha \geq 2$, the sum defining $H_\alpha(P)$ runs over all $i_1, \ldots, i_{\alpha-1}, i_\alpha$ for which $p(x_{i_1}, \ldots, x_{i_{\alpha-1}}) > 0$. $H_\alpha(P)$ is non-negative since $0 \leq p(x_{i_1}) \leq 1$ and $0 \leq p(x_{i_\alpha} | x_{i_1}, \ldots, x_{i_{\alpha-1}}) \leq 1$. $H_\alpha(P)$ is known as the *conditional entropy of* X_α *given* $X_1, \ldots, X_{\alpha-1}$. As in Lemma IX.2.1(b),

$$I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P) = -H_1(P) - H_2(P) - \cdots - H_\alpha(P) - \alpha \sum_{i_1=1}^{r} p(x_{i_1}) \log \rho_{i_1}$$

and $H_1(P) \geq H_2(P) \geq \cdots \geq H_\alpha(P)$. Since $H_\alpha(P)$ is non-negative, the non-increasing sequence $H_1(P), H_2(P), \ldots$ has a limit which we denote by $h(P)$ (the *mean entropy* of P). It follows that $\lim_{\alpha\to\infty} \alpha^{-1} I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P)$ exists and

$$\lim_{\alpha\to\infty} \frac{1}{\alpha} I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P) = -h(P) - \sum_{i_1=1}^{r} p(x_{i_1}) \log \rho_{i_1} < \infty.$$

The existence of $\lim_{\alpha\to\infty} \alpha^{-1} I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P)$ is also a consequence of the super-additivity of the sequence $\{I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P); \alpha = 1, 2, \ldots\}$ [page 274] and the bounds

$$0 \leq I_{\pi_\alpha P_\rho}^{(2)}(\pi_\alpha P) \leq \alpha \sum_{i=1}^{r} \log \frac{1}{\rho_i}, \qquad \alpha = 1, 2, \ldots.$$

IX.6. Problems

IX.6.1. Consider the function $I_{\rho,\alpha}^{(3)}(\tau)$, $\alpha \geq 2$, defined in (9.6). Using Theorem IX.3.3 (contraction principle) and the fact that $I_{\rho}^{(3)}(P)$ is lower semicontinuous and affine [Section IX.2], prove that $I_{\rho,\alpha}^{(3)}(\tau)$ is a closed convex function on $\mathcal{M}_s(\Gamma^\alpha)$.

IX.6.2. Here is another proof of the convexity of $I_{\rho,\alpha}^{(3)}(\tau)$, $\tau \in \mathcal{M}_s(\Gamma^\alpha)$. $\mathcal{M}_s(\Gamma^\alpha)$ is homeomorphic to the compact convex subset $\mathcal{M}_{s,\alpha}$ of $\mathbb{R}^{r^\alpha}$ defined after the proof of Lemma IX.4.1. By (9.13) and (9.19),

$$I_{\rho,\alpha}^{(3)}(\tau) = \bar{I}_{\rho,\alpha}^{(3)}(\tau) = \sup_{t \in \mathbb{R}^{r^\alpha}} \{\langle t, \tau \rangle - \log \lambda(B_\alpha(t))\} \qquad \text{for } \tau \in \mathcal{M}_{s,\alpha}.$$

For $\tau \in \mathbb{R}^{r^\alpha} \backslash \mathcal{M}_{s,\alpha}$, $\bar{I}_{\rho,\alpha}^{(3)}(\tau)$ is defined to be ∞.

(a) Using the last display, prove that $\bar{I}_{\rho,\alpha}^{(3)}$ is a closed convex function on $\mathbb{R}^{r^\alpha}$. It follows that $I_{\rho,\alpha}^{(3)}$ is a closed convex function on $\mathcal{M}_s(\Gamma^\alpha)$.

(b) Prove that $\bar{I}_{\rho,\alpha}^{(3)}$ is essentially strictly convex. [*Hint:* Theorem VI.5.6.]

IX.6.3. (a) Prove that for any $\tau \in \mathcal{M}_s(\Gamma^\alpha)$, $\alpha \geq 2$, $I_{\rho,\alpha}^{(3)}(\tau)$ is non-negative and $I_{\rho,\alpha}^{(3)}(\tau)$ equals 0 if and only if $\tau = \pi_\alpha P_\rho$ (finite product measure).

(b) Let ν be a probability measure on $\mathscr{B}(\Gamma)$. Prove that $I_{\rho,\alpha}^{(3)}(\tau)$ attains its infimum over the set $\{\tau \in \mathcal{M}_s(\Gamma^\alpha) : \pi_1 \tau = \nu\}$ at the unique measure $\pi_\alpha P_\nu$ and

$$\inf\{I_{\rho,\alpha}^{(3)}(\tau) : \tau \in \mathcal{M}_s(\Gamma^\alpha), \pi_1 \tau = \nu\} = I_{\rho,\alpha}^{(3)}(\pi_\alpha P_\nu) = I_\rho^{(2)}(\nu).$$

IX.6.4. Let $B = \{B_{ij}\}$ be a positive $m \times m$ matrix (some $m \geq 2$). Let u and w be positive left and right eigenvectors associated with $\lambda(B)$ and normalized so that $\langle u, w \rangle = 1$. Denote by $\mathcal{M}_{s,2}$ the set of all $\tau = \{\tau_{ij}; i,j = 1, \ldots, m\}$ which satisfy $\tau_{ij} \geq 0$ for each i and j, $\sum_{i,j=1}^m \tau_{ij} = 1$, and $\sum_{j=1}^m \tau_{ij} = \sum_{k=1}^m \tau_{ki}$ for each i. Let $(\nu_\tau)_i = \sum_{j=1}^m \tau_{ij}$. Prove that

$$\log \lambda(B) = \sup_{\tau \in \mathcal{M}_{s,2}} \sum \tau_{ij} \log \frac{B_{ij}(\nu_\tau)_i}{\tau_{ij}}$$

and that the supremum is attained at the unique point $\tau_{ij}^0 = u_i B_{ij} w_j / \lambda(B)$. The sum in the last display runs over all i and j for which $(\nu_\tau)_i > 0$. [*Hint:* τ^0 is the unique point in ri $\mathcal{M}_{s,2}$ at which $f(\tau) = \sum_{i,j=1}^m \tau_{ij} \log [B_{ij}(\nu_\tau)_i / \tau_{ij}]$ attains its supremum [Theorem IX.4.4]. If $\rho_1, \ldots, \rho_m$ are positive numbers which sum to 1, then

$$f(\tau) = \sum_{i,j=1}^m \tau_{ij} \log(B_{ij}/\rho_j) - I_{\rho,2}^{(3)}(\tau).$$

Complete the proof using Problem IX.6.1.]

The remaining problems concern finite-state Markov chains. We use the following notation.

Γ a finite set $\{x_1, x_2, \ldots, x_r\}$ with $x_1 < x_2 < \cdots < x_r$.

$\gamma = \{\gamma_{ij}\}$ a positive $r \times r$ stochastic matrix.

$\nu = \{\nu_i; i = 1, \ldots, r\}$ the unique positive solution of the equations $\sum_{i=1}^r \nu_i \gamma_{ij} = \nu_j$, $\sum_{i=1}^r \nu_i = 1$.

P_γ the Markov chain in $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ with transition matrix γ and invariant measure ν.

$\{X_j; j \in \mathbb{Z}\}$ the coordinate representation process on $\Gamma^{\mathbb{Z}}$.

$\lambda(B)$ the largest positive eigenvalue of a primitive matrix B [Lemma IX.4.1].

$\mathscr{M}$ the set $\{\sigma \in \mathbb{R}^r : \sigma_i \geq 0, \sum_{i=1}^r \sigma_i = 1\}$. $\mathscr{M}$ is homeomorphic to the set $\mathscr{M}(\Gamma)$ of probability measures on $\mathscr{B}(\Gamma)$; $\sigma \in \mathscr{M}$ corresponds to the measure on $\mathscr{B}(\Gamma)$ with $\sigma\{x_i\} = \sigma_i$.

IX.6.5. [Renyi (1970a), Spitzer (1971)]. The purpose of this problem is to prove the limit $\lim_{n\to\infty} \gamma_{ij}^n = \nu_j$ using relative entropy [see Lemma A.9.5.]

(a) Using Lemma IX.4.1, prove that the equations $\sum_{i=1}^r \nu_i \gamma_{ij} = \nu_j$, $\sum_{i=1}^r \nu_i = 1$ have a unique positive solution $\nu = \{\nu_i; i = 1, \ldots, r\}$.

(b) Given $\sigma \in \mathscr{M}$, define $\sigma\gamma^n \in \mathscr{M}$ by $(\sigma\gamma^n)_j = \sum_{i=1}^r \sigma_i \gamma_{ij}^n$. Set $J_n(\sigma) = \sum_{i=1}^r \nu_i \log(\nu_i/(\sigma\gamma^n)_i)$. Prove that for each $j \in \{1, \ldots, r\}$,

$$\log \frac{\nu_j}{(\sigma\gamma^{n+1})_j} \leq \sum_{i=1}^r \nu_i \frac{\gamma_{ij}}{\nu_j} \log \frac{\nu_i}{(\sigma\gamma^n)_i}$$

with equality iff $(\sigma\gamma^n)_i = \nu_i$. Deduce that $J_{n+1}(\sigma) \leq J_n(\sigma)$ with equality iff $\sigma\gamma^n = \nu$.

(c) Let $\{n'\}$ be a subsequence such that $\bar{\nu} = \lim_{n'\to\infty} \sigma\gamma^{n'}$ exists. Prove that $\lim_{n'\to\infty} J_{n'+n}(\nu) = J_n(\bar{\nu})$ for all $n \in \{1, 2, \ldots\}$. Deduce that $\sigma\gamma^n \to \nu$ for any $\sigma \in \mathscr{M}$. It follows that $\lim_{n\to\infty} \gamma_{ij}^n = \nu_j$ for each i and j.

IX.6.6 [Ellis (1981)]. Part (a) proves the level-1 large deviation property.

(a) Let $B(t)$ be the $r \times r$ matrix $\{e^{tx_i}\gamma_{ij}; i,j = 1, \ldots, r\}$, $t \in \mathbb{R}$. Prove that the P_γ-distributions of $S_n/n = \sum_{j=1}^n X_j/n$ have a large deviation property with $a_n = n$ and entropy function

$$I_\gamma^{(1)}(z) = \sup_{t \in \mathbb{R}} \{tz - \log \lambda(B(t))\} \qquad \text{for } z \in \mathbb{R}.$$

(b) Prove that $I_\gamma^{(1)}(z) \geq 0$ with equality if and only if z equals $z_0 = \sum_{i=1}^r x_i \nu_i$.

(c) Prove that $S_n/n \xrightarrow{\exp} \sum_{i=1}^r x_i \nu_i$.

IX.6.7. This problem proves the level-2 large deviation property.

(a) Let $L_n(\omega, \cdot) = n^{-1} \sum_{j=1}^n \delta_{X_j(\omega)}(\cdot)$, $n = 1, 2, \ldots$ (empirical measure). Prove that $L_n(\omega, \cdot) \Rightarrow \nu$ P_γ-a.s.

(b) [Ellis (1984)]. Let $B(t)$ be the $r \times r$ matrix $\{e^{t_i}\gamma_{ij}; i,j = 1, \ldots, r\}$, $t \in \mathbb{R}^r$. Prove that the P_γ-distributions of $\{L_n; n = 1, 2, \ldots\}$ on $\mathscr{M}(\Gamma)$ have a large deviation property with $a_n = n$ and entropy function

$$I_\gamma^{(2)}(\sigma) = \sup_{t \in \mathbb{R}^r} \{\langle t, \sigma\rangle - \log \lambda(B(t))\} \qquad \text{for } \sigma \in \mathscr{M}(\Gamma).$$

(c) Let $\mathcal{N}^+$ denote the set of vectors $t \in \mathbb{R}^r$ of the form $t_i = \log[u_i/(\gamma u)_i]$ for some $u > 0$ in $\mathbb{R}^r$ ($(\gamma u)_i = \sum_{j=1}^r \gamma_{ij} u_j$). Prove that $\lambda(B(t)) = 1$ if and only if $t \in \mathcal{N}^+$ and that for any $t \in \mathbb{R}^r$

$$h(t) = t - [\log \lambda(B(t))]\mathbf{1} \text{ belongs to } \mathcal{N}^+ \ (\mathbf{1} = (1, \ldots, 1)).$$

(d) [Donsker and Varadhan (1975a)]. Prove that for $\sigma \in \mathcal{M}(\Gamma)$, $I_\gamma^{(2)}(\sigma) = -\inf_{u>0} \sum_{i=1}^r \sigma_i \log[(\gamma u)_i/u_i]$.

(e) Prove that $I_\gamma^{(2)}(\sigma) \geq 0$ with equality if and only if $\sigma = \nu$.

IX.6.8. Let B be a positive $r \times r$ matrix. The purpose of this problem is to prove that

$$\lambda(B) = \sup_{\sigma \in \mathcal{M}} \inf_{u>0} \sum_{i=1}^r \sigma_i \frac{(Bu)_i}{u_i}. \tag{9.26}$$

For generalizations, see Donsker and Varadhan (1975e), Friedland and Karlin (1975), and Friedland (1981).

(a) Let $B(t) = \{e^{t_i}\gamma_{ij}; i,j = 1, \ldots, r\}$, $t \in \mathbb{R}^r$, where $\gamma = \{\gamma_{ij}; i,j = 1, \ldots, r\}$ is a positive stochastic matrix. Prove that

$$\log \lambda(B(t)) = \sup_{\sigma \in \mathcal{M}} \left\{ \langle t, \sigma \rangle + \inf_{u>0} \sum_{i=1}^r \sigma_i \log \frac{(\gamma u)_i}{u_i} \right\} \qquad \text{for } t \in \mathbb{R}^r.$$

(b) Prove that $\log \lambda(B) = \sup_{\sigma \in \mathcal{M}} \inf_{u>0} \sum_{i=1}^r \sigma_i \log[(Bu)_i/u_i]$ by suitable choice of t and γ in part (a). Derive (9.26).

IX.6.9 [Kac (1980)]. Let B be a positive $r \times r$ matrix which is also symmetric ($B_{ij} = B_{ji}$). Show that (9.26) reduces to the Rayleigh–Ritz formula

$$\lambda(B) = \sup \left\{ \sum_{i,j=1}^r B_{ij} v_i v_j : v_i \text{ real}, \sum_{i=1}^r v_i^2 = 1 \right\}.$$

[*Hint:* If $\sigma \in \mathcal{M}$ is positive, then with $w_i = u_i/\sqrt{\sigma_i}$ $(u > 0)$

$$\sum_{i=1}^r \sigma_i \frac{(Bu)_i}{u_i} = \frac{1}{2} \sum_{i,j=1}^r B_{ij} \sqrt{\sigma_i} \sqrt{\sigma_j} \left(\frac{w_i}{w_j} + \frac{w_j}{w_i} \right).]$$

The next six problems show how to prove the level-3 large deviation property.

IX.6.10. For $P \in \mathcal{M}_s(\Gamma^{\mathbb{Z}})$, set $I_{\pi_\alpha P_\gamma}^{(2)}(\pi_\alpha P) = \sum_{\omega \in \Gamma^\alpha} \pi_\alpha P\{\omega\} \log[\pi_\alpha P\{\omega\} / \pi_\alpha P_\gamma\{\omega\}]$. Prove that $I_\gamma^{(3)}(P) = \lim_{\alpha \to \infty} \alpha^{-1} I_{\pi_\alpha P_\gamma}^{(2)}(\pi_\alpha P)$ exists and express $I_\gamma^{(3)}(P)$ as in (9.1).

IX.6.11. (a) Prove that $I_\gamma^{(3)}(P)$ is lower semicontinuous, has compact level sets, and is affine.

(b) Prove that $I_\gamma^{(3)}(P) \geq 0$ and that equality holds if and only if $P = P_\gamma$. [*Hint:* Problems IX.6.7(e) and IX.6.13.]

IX.6.12. For $\tau \in \mathscr{M}_s(\Gamma^\alpha)$, $\alpha \geq 2$, set $\tau_{i_1 \cdots i_\alpha} = \tau\{x_{i_1}, \ldots, x_{i_\alpha}\}$ and $(v_\tau)_{i_1, \ldots, i_{\alpha-1}} = \sum_{i_\alpha=1}^r \tau_{i_1 \cdots i_\alpha}$ and define

$$I_{\gamma,\alpha}^{(3)}(\tau) = \sum \tau_{i_1 \cdots i_\alpha} \log \frac{\tau_{i_1 \cdots i_\alpha}}{(v_\tau)_{i_1 \cdots i_{\alpha-1}} \gamma_{i_{\alpha-1} i_\alpha}},$$

where the sum runs over all $i_1, \ldots, i_{\alpha-1}, i_\alpha$ for which $(v_\tau)_{i_1 \cdots i_{\alpha-1}} > 0$. Prove that

$$\inf\{I_\gamma^{(3)}(P): P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \pi_\alpha P = \tau\} = I_{\gamma,\alpha}^{(3)}(\tau) \tag{9.27}$$

and determine the measure(s) P at which the infimum is attained. Formula (9.27) is a contraction principle relating $I_\gamma^{(3)}$ and $I_{\gamma,\alpha}^{(3)}$.

IX.6.13. The level-2 entropy function is given by $I_\gamma^{(2)}(\sigma) = -\inf_{u>0} \sum_{i=1}^r \sigma_i \log[(\gamma u)_i/u_i]$, $\sigma \in \mathscr{M}(\Gamma)$ [Problem IX.6.7]. The contraction principle relating $I_\gamma^{(3)}$ and $I_\gamma^{(2)}$ states that

$$\inf\{I_\gamma^{(3)}(P): P \in \mathscr{M}_s(\Gamma^{\mathbb{Z}}), \pi_1 P = \sigma\} = I_\gamma^{(2)}(\sigma) \qquad \text{for } \sigma \in \mathscr{M}(\Gamma). \tag{9.28}$$

According to (9.27), (9.28) can be proved by showing that

$$\inf\{I_{\gamma,2}^{(3)}(\tau): \tau \in \mathscr{M}_s(\Gamma^2), \pi_1 \tau = \sigma\} = I_\gamma^{(2)}(\sigma) \qquad \text{for } \sigma \in \mathscr{M}(\Gamma). \tag{9.29}$$

Note that $(\pi_1 \tau)_i = (v_\tau)_i = \sum_{j=1}^r \tau_{ij}$. Prove (9.29) and determine the measure(s) P at which the infimum in (9.28) is attained.
[*Hint:* For $\sigma > 0$ and $u > 0$, let $f_\sigma(u) = \sum_{i=1}^r \sigma_i \log[(\gamma u)_i/u_i]$. Show that if $\xi > 0$ is a minimum point of f_σ, then $I_{\gamma,2}^{(3)}(\tau)$ attains its infimum over the set $\{\tau \in \mathscr{M}_s(\Gamma^2): v_\tau = \sigma\}$ at the unique measure $\tau_{ij} = \sigma_i \gamma_{ij} \xi_j/(\gamma\xi)_i$.]

IX.6.14. Let $R_n(\omega, \cdot) = n^{-1} \sum_{k=0}^{n-1} \delta_{T^k X(n,\omega)}(\cdot)$, $n = 1, 2, \ldots$ (empirical process). For $\alpha \geq 2$, prove that the P_γ-distributions of the α-dimensional marginals $\{\pi_\alpha R_n(\omega, \cdot); n = 1, 2, \ldots\}$ on $\mathscr{M}_s(\Gamma^\alpha)$ have a large deviation property with $a_n = n$ and entropy function $I_{\gamma,\alpha}^{(3)}$.

IX.6.15. (a) Prove that $R_n(\omega, \cdot) \Rightarrow P_\gamma$ P_γ-a.s.

(b) Prove that the P_γ-distributions of $\{R_n(\omega, \cdot)\}$ on $\mathscr{M}_s(\Gamma^{\mathbb{Z}})$ have a large deviation property with $a_n = n$ and entropy function $I_\gamma^{(3)}$.