The Annals of Probability
1993, Vol. 21, No. 1, 216–231

IDENTIFYING A LARGE DEVIATION RATE FUNCTION

BY I. H. DINWOODIE

Tulane University

Assume a sequence of probabilities $\{P_n\}$ has a large deviation rate function I. It is proved that I takes a form analogous to a convex conjugate. If I is also assumed convex, then I is a convex conjugate of an explicitly defined function ψ. The results are applied to the empirical law of a Markov chain yielding universal bounds on I. Examples are given of Markov chains in which the empirical law has a large deviation rate strictly between the given bounds.

1. Introduction. Suppose a sequence of measures $\{P_n\}$ on a Polish space (X, d) has a large deviation rate function I: $X \to [0, \infty]$, by which we mean that I is lower semicontinuous and

$$\liminf n^{-1} \log P_n(U) \geq -I(U), \tag{1.1}$$

$$\limsup n^{-1} \log P_n(C) \leq -I(C) \tag{1.2}$$

for all open sets U and closed sets C contained in X. We are using the customary notation in which $I(A) = \inf\{I(x): x \in A\}$ whenever $A \subset X$. We will say that a function I is a *lower rate function* if (1.1) holds for all open sets or an *upper rate function* if (1.2) holds for all closed sets. This article seeks the form of a large deviation rate function I and establishes certain inevitable formalities under the assumption that I exists. It is a basic and well-known fact that a large deviation rate function I is unique (see Lemma 1.1 below). By starting with a generalization of a theorem of Varadhan (1966, 1984), we identify in Theorem 2.1 the rate function in terms of a quantity analogous to a convex conjugate. This identification is valid regardless of whether the rate function is convex or not.

In Section 3 the rate function I is assumed convex and a simpler representation for I is given in terms of a convex conjugate. In Section 4 we apply the results of Section 3 to the empirical probability measure L_n in the weak topology of a Markov chain on a Polish state space S. Theorem 4.1 gives a precise characterization of a convex large deviation rate function in terms of a convex conjugate. Theorem 4.2 shows that a convex rate function always lies between the convex conjugate of the logarithm of the reciprocal of the convergence parameter of a certain irreducible kernel T_ξ and the convex conjugate of the logarithm of the spectral radius of the same kernel considered as an operator on the continuous bounded functions on the state space of the chain. Example 4.1 shows that the correct rate function may be strictly between

Received October 1990; revised June 1991.
AMS 1991 *subject classification*. Primary 60F10.
Key words and phrases. Large deviations, Markov chain.

these two rates, depending on the initial distribution of the Markov chain. Finally, Proposition 4.1 shows that for initial distributions concentrated on a certain core of the state space, which we call a minimal closed set, the empirical law can only have the convex conjugate of the logarithm of the reciprocal of the convergence parameter as a rate function. The results of Section 4 complement and to some extent are based on results of de Acosta (1985, 1988, 1990) and Ney and Nummelin (1987).

We begin with two well-known results for completeness.

LEMMA 1.1. *A large deviation rate function is unique.*

PROOF. Suppose that I and J are two rate functions for a sequence of probabilities $\{P_n\}$. If I and J are different, then there is a point where I or J is superior to the other. We can assume that $I(x) > J(x)$. Let $c > 0$ be such that

$$0 \le J(x) < c < I(x) \le \infty.$$

By the lower semicontinuity of I, there exists $\varepsilon > 0$ such that if $y \in \overline{B}_{x,\varepsilon} = \{z\colon d(z,x) \le \varepsilon\}$, then $I(y) > c$. Let $B_{x,\varepsilon} = \{z\colon d(z,x) < \varepsilon\}$. Then, by the large deviation property,

$$\begin{aligned} -J(x) &\le -J(B_{x,\varepsilon}) \\ &\le \liminf \frac{1}{n} \log P_n(B_{x,\varepsilon}) \\ &\le \limsup \frac{1}{n} \log P_n(\overline{B}_{x,\varepsilon}) \\ &\le -I(\overline{B}_{x,\varepsilon}) \\ &\le -c, \end{aligned}$$

which contradicts the fact that $J(x) < c$. □

REMARK 1.1. The uniqueness of the rate function is proved as Theorem II.3.2 in Ellis (1985) under the additional assumption that I have compact level sets. Ellis credits the result to Varadhan. The result is stated in Orey [(1986), page 202] without the hypothesis of compactness, but the proof presented here seems to be the simplest.

Recall the following slightly generalized version of a theorem of Varadhan (1966, 1984).

THEOREM 1.1. *Let $\{P_n\}$ satisfy the large deviation principle with rate function I. Then for any continuous function $f\colon X \to \mathbf{R}$ that is bounded above,*

$$\lim n^{-1} \log \int_X e^{nf(x)}\, dP_n = -\inf_{x \in X} \{I(x) - f(x)\}.$$

The theorem differs from that of Varadhan in allowing f to be unbounded below. Varadhan's proof generalizes with no difficulties.

2. General rate functions. Theorem 1.1 can be used to describe the rate function I in a precise way whenever it exists. Let $\mathbf{C}(X)$ denote the bounded continuous functions on X, and let

$$\phi(f) = \lim n^{-1} \log \int_X e^{n f(x)}\, dP_n$$

for $f \in \mathbf{C}(X)$. Theorem 2.1 below shows that ϕ: $\mathbf{C}(X) \to [-\infty, \infty)$ contains all the large deviation information for $\{P_n\}$ and confirms the importance of the formalism of the convex conjugate in large deviations. The idea of characterizing the rate function I through a version of Theorem 1.1 seems to have first appeared in Stroock (1984) for the special case of the occupation time for Markov chains. Similar ideas appear in Bryc (1990). The function ϕ could be called the free energy function [see Ellis (1985), page 46], or it could be called the pressure of the sequence $\{P_n\}$ by analogy with the topological pressure of Ruelle [(1973), Theorem 5.1].

THEOREM 2.1. *If $\{P_n\}$ satisfies the large deviation principle with rate function I, then*

$$I(x) = \sup_{f \in \mathbf{C}(X)} \{f(x) - \phi(f)\}.$$

PROOF. Since $\phi(f) = \sup_y\{f(y) - I(y)\}$, it is immediate that

$$I(x) \geq \sup_{f \in \mathbf{C}(X)} \{f(x) - \phi(f)\}.$$

To show that $I(x) \leq \sup_{f \in \mathbf{C}(X)}\{f(x) - \phi(f)\}$, we can assume that $I(x) > 0$. Let $c < I(x)$ and choose $\varepsilon > 0$, using the lower semicontinuity of I, such that $I(y) > c$ for each $y \in \overline{B}_{x, 2\varepsilon} = \{y\colon d(y, x) \leq 2\varepsilon\}$. Fix $M \geq 0$, and let f_M be a continuous, bounded function defined on X such that

$$f_M(y) = 0, \qquad d(y,x) \leq \varepsilon,$$
$$f_M(y) = -M, \qquad d(y,x) \geq 2\varepsilon,$$

and $-M \leq f_M(y) \leq 0$ for $\varepsilon < d(y, x) < 2\varepsilon$. Now

$$\begin{aligned}\phi(f_M) &= \lim \frac{1}{n} \log\left[\int_{d(y,x)<2\varepsilon} \exp(nf_M)\, dP_n + \int_{d(y,x)\geq 2\varepsilon} \exp(nf_M)\, dP_n\right]\\ &\leq \limsup \frac{1}{n} \log\left[P_n\left(\overline{B}_{x,2\varepsilon}\right) + e^{-nM}\right]\\ &\leq \max\left\{-I\left(\overline{B}_{x,2\varepsilon}\right), -M\right\}\\ &= -\min\left\{I\left(\overline{B}_{x,2\varepsilon}\right), M\right\},\end{aligned}$$

and therefore $\min\{I(\overline{B}_{x,2\varepsilon}), M\} \leq -\phi(f_M) = f_M(x) - \phi(f_M)$. Now if

$I(\overline{B}_{x,2\varepsilon}) = \infty$, this would imply that $M \le f_M(x) - \phi(f_M)$, and hence

$$I(x) = \infty = \sup_{f \in \mathbf{C}(X)} \{f(x) - \phi(f)\}.$$

Otherwise, choose $M > I(\overline{B}_{x,2\varepsilon})$ so that

$$c < I(\overline{B}_{x,2\varepsilon}) \le f_M(x) - \phi(f_M) \le \sup_{f \in \mathbf{C}(X)} \{f(x) - \phi(f)\}.$$

Now since c was chosen arbitrarily such that $0 < c < I(x) \le \infty$, it follows that

$$I(x) \le \sup_{f \in \mathbf{C}(X)} \{f(x) - \phi(f)\},$$

which completes the proof. □

Theorem 2.1 can be used when a rate function may not be convex. For example, an exchangeable sequence of 0-1 valued random variables has a nonconvex rate function related to the de Finetti decomposition [Ellis (1984), page 3, Dinwoodie and Zabell (1992)]. A Markov mixture example is given below.

EXAMPLE 2.1. Let the transition matrix for a Markov chain on $\{1, 2, 3, 4\}$ be given by

$$\pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let the initial distribution be $(\alpha, 0, 1, -\alpha, 0)$ for $0 < \alpha < 1$. Then the law P_n of the empirical probability measure L_n is the convex combination

$$p_n = [\alpha] P_1 \circ L_n^{-1} + [1 - \alpha] P_3 \circ L_n^{-1},$$

where P_1 is a Markov measure on $\{1, 2, 3, 4\}^\infty$ with initial distribution $(1, 0, 0, 0)$ and transition matrix π and P_3 is a Markov measure on $\{1, 2, 3, 4\}^\infty$ with initial distribution $(0, 0, 1, 0)$ and transition matrix π. It is not hard to see that $P_1 \circ L_n^{-1}$, $i = 1, 3$, has large deviation rate function λ_i on the probability simplex in $\mathbf{R}^4$ given by

$$\lambda_1(p_1, p_2, p_3, p_4) = \begin{cases} p_1 \log 2, & \text{if } p_3 = p_4 = 0, \\ \infty, & \text{otherwise}; \end{cases}$$

$$\lambda_3(p_1, p_2, p_3, p_4) = \begin{cases} p_3 \log 2, & \text{if } p_1 = p_2 = 0, \\ \infty, & \text{otherwise}. \end{cases}$$

Now from Theorem 2.1, if $\{P_n\}$ has a large deviation rate it must be

$$\begin{aligned} I(x) &= \sup_f \{f(x) - \phi(f)\} \\ &= \sup_f \left\{ f(x) - \max_{i=1,3} \left\{ \sup_y [f(y) - \lambda_i(y)] \right\} \right\} \\ &= \sup_f \min_{i=1,3} \inf_y \{f(x) - f(y) + \lambda_i(y)\}. \end{aligned}$$

Then $I(x)$ is inferior to $\min_{i=1,3} \lambda_i(x)$, which can be seen by setting $y = x$ in the infimum of the above expression. But by setting f to be $\min_{i=1,3} \lambda_i(x)$ in the supremum of the above expression, $I(x)$ is also seen to be superior to $\min_{i=1,3} \lambda_i(x)$. Hence the large deviation rate function must be $\min_{i=1,3} \lambda_i(x)$ if such a rate exists. It is easy to see that $\min_{i=1,3} \lambda_i(x)$ is in fact the large deviation rate function for $\{P_n\}$.

The empirical probability measure for the following Markov chain also has a nonconvex rate function. The Markov chain is irreducible and ergodic according to the general theory of Markov chains developed in Nummelin (1984).

EXAMPLE 2.2. Let the transition matrix for a Markov chain on $\{1, 2, 3, 4\}$ be given by

$$\pi = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let the initial distribution be δ_1 and let $L_n = (L_{n,j})$, $1 \le j \le 4$, denote the empirical law of the Markov chain as an element of the probability simplex in $\mathbf{R}^4$, where

$$L_{n,j} = \frac{1}{n} \sum_{i=1}^{n} 1 \circ \{X_i = j\}.$$

If the law of the sequence L_n had a large deviation rate function I, we could find I using Theorem 2.1. For f defined on the probability simplex in $\mathbf{R}^4$,

$$\begin{aligned} \phi(f) = \max\Big\{ &\sup_\alpha \{f(0,0,\alpha,1-\alpha) - \alpha \log 2\}, \\ &\sup_\alpha \{f(0,\alpha,0,1-\alpha) - \alpha \log 2\} \Big\}, \end{aligned}$$

where $0 \le \alpha \le 1$. Now if $p = (p_i)$ belongs to the probability simplex in $\mathbf{R}^4$ and $p_1 \gtrless 0$, let $f_c(x_1, x_2, x_3, x_4) = cx_1$ to see that $I(p) \ge f_c(p) - \phi(f_c) = cp_1$, for any $c > 0$, and thus

$$I(p) = \infty \quad \text{if } p_1 > 0.$$

Similarly, one can use the function $f_c(x_1, x_2, x_3, x_4) = c(x_2 \wedge x_3)$, $c > 0$, to see that

$$I(p) = \infty \quad \text{if } p_2 > 0 \text{ and } p_3 > 0.$$

Finally, excluding these two cases, one can show that

$$I(p) = (1 - p_4)\log 2 \quad \text{otherwise}.$$

In fact $P \circ L_n^{-1}$ has large deviation rate function I on the probability simplex in $\mathbf{R}^4$.

REMARK 2.1. A probabilistic characterization of the rate function follows immediately from the definition. Let $B_{x,\varepsilon}$ denote the open ball of radius $\varepsilon > 0$ centered at $x \in X$ and let $\overline{B}_{x,\varepsilon}$ be its closure. If $\{P_n\}$ satisfies the large deviation principle with rate function I, then

$$\begin{aligned} I(x) &= -\lim_{\varepsilon \to 0} \limsup_{n} \frac{1}{n} \log P_n(\overline{B}_{x,\varepsilon}) \\ &= -\lim_{\varepsilon \to 0} \liminf_{n} \frac{1}{n} \log P_n(B_{x,\varepsilon}). \end{aligned}$$

This is immediate since the lower semicontinuity of I implies

$$I(x) = \lim_{\varepsilon \to 0} I(B_{x,\varepsilon}) = \lim_{\varepsilon \to 0} I(\overline{B}_{x,\varepsilon}).$$

3. Convex rate functions. The goal is to replace $\mathbf{C}(X)$ with a smaller class of functions which could simplify the calculation of I in the representation of Theorem 2.1. Assume henceforth that X is a closed, convex subset of a locally convex and separated space E whose relative topology is that of a Polish space. Extend I to E by setting $I(v) = \infty$ when $v \in E - X$. The function I extended to E remains a large deviation rate function for the sequence $\{P_n\}$ extended to E. We try to replace $\mathbf{C}(X)$ with the continuous linear functions on E, which we denote by E'. The function I must be convex if substituting E' for $\mathbf{C}(X)$ is to yield I.

If $f: E \to (-\infty, \infty]$, let $f^*: E' \to (-\infty, \infty]$ denote its convex conjugate given by

$$f^*(\xi) = \sup_{v \in E} \langle v, \xi \rangle - f(v).$$

It is a basic property of the convex conjugate that if $f: E \to (-\infty, \infty]$ is lower semicontinuous and convex, then $f^{**} = f$ when the dual of E' in the second transformation on f^* is taken to be E [see Ekeland and Temam (1974), Proposition 4.1].

Define $\psi(\xi)$ for $\xi \in E'$ by

$$\psi(\xi) = \lim_{M \to \infty} \phi(\langle \cdot , \xi \rangle \wedge M),$$

which exists according to Theorem 1.1.

THEOREM 3.1. *Suppose that I is a convex large deviation rate function for the sequence $\{P_n\}$. Then $\psi^* = I$.*

PROOF. If I is convex and l.s.c., then $I = I^{**}$. The theorem will follow if we can show that $\psi = I^*$. Now

$$\begin{aligned}\psi(\xi) &= \lim_{M\to\infty} \phi(\langle \cdot , \xi\rangle \wedge M)\\ &= \lim_{M\to\infty} \sup_{x\in X} [\langle x, \xi\rangle \wedge M - I(x)]\\ &= \sup_M \sup_{x\in X} [\langle x, \xi\rangle \wedge M - I(x)]\\ &= \sup_{x\in X} [\langle x, \xi\rangle - I(x)]\\ &= \sup_{v\in E} [\langle v, \xi\rangle - I(v)]\\ &= I^*(\xi),\end{aligned}$$

which proves the result. □

Let $\bar{\psi}(\xi) = \lim\sup(1/n)\log E_n \exp\langle \cdot , \xi\rangle$. Many large deviation results concern themselves with $\bar{\psi}^*$ rather than ψ^* [Gärtner (1977), Ellis (1984), de Acosta (1985)]. In particular, it is a result of de Acosta [(1985), Theorem 2.1] that $\bar{\psi}^*$ always works as an upper bound for compact sets whether the sequence $\{P_n\}$ has or does not have a large deviation rate I. It is immediate from the definitions that

$$\psi(\xi) \le \bar{\psi}(\xi)$$

and thus we have the following corollary of Theorem 3.1.

COROLLARY 3.1. *If the sequence of probabilities $\{P_n\}$ has a convex large deviation rate function I, then $\bar{\psi}^*$ is an upper rate function for all closed sets.*

PROOF. $I = \psi^* \ge \bar{\psi}^*$. □

Suppose for the moment that P_n is the law of the mean of an i.i.d. sequence of random vectors $\{X_i\colon i \ge 1\}$. From the monotone convergence theorem it follows that

$$\bar{\psi}(\xi) = \log Ee^{\xi(X_1)} = \psi(\xi).$$

Thus, whenever the law of the mean of an i.i.d. sequence of random vectors has a large deviation rate function, the rate function must be the convex conjugate of the logarithm of the moment generating function. Conditions under which this sequence of probabilities does in fact have a large deviation rate function can be found in Bahadur and Zabell (1979). In particular, it follows easily from their results that the mean $\bar{X}_n$ of any i.i.d. sequence $\{X_i\colon i \ge 1\}$ in $\mathbf{R}^1$ does indeed have a convex large deviation rate function. Then

Theorem 3.1 can be used to give a simplified if indirect proof of Chernoff's theorem [Chernoff (1952), Theorem 1]. Most proofs of this fundamental result become complicated when dealing rigorously with the case where the moment generating function of X_1 does not exist [see Bahadur (1971), Theorem 3.1, Azencott (1980), Théorème 2.3]. Chernoff credits Cramér (1938) with the result under strong conditions, but Chernoff furnished in the general case only what he himself described as a brief sketch of a proof.

REMARK 3.1. The assumption that $\{P_n\}$ has a convex rate function I also permits a strengthening of the probabilistic characterization of I. If $\{U_\alpha\}_{\alpha \in A}$ is any base of open and convex sets at $x \in E$, then when $I(x) < \infty$,

$$I(x) = \sup_\alpha \{I(U_\alpha)\}, \tag{3.1}$$

$$\lim \frac{1}{n} \log P_n(U_\alpha) = -I(U_\alpha). \tag{3.2}$$

Statement (3.1) follows from Theorem 2.2 and the local convexity of E. Statement (3.2) follows from the fact that the segment joining x to any boundary point of U_α is contained in U_α and thus when $I(x) < \infty$ and I is convex, $I(U_\alpha) = I(\overline{U}_\alpha)$. Example 2.1 shows that (3.1) and (3.2) may hold even when I is not convex.

In Section 4 we proceed with computing ψ and ψ^* for the empirical law L_n of a Markov chain.

4. Application to Markov chains. In this section we turn to computing $\psi(\xi)$ when studying the empirical probability measure L_n of a Markov chain. Let S be a Polish space with Borel field B and let $\{\pi(x, \cdot\,): x \in S\}$ be a family of transition probabilities on S. Let μ be an initial distribution on S and define the Markov probability P_μ on the pair $(\Omega = S^\infty, B(S^\infty))$ so that for $A_1 \times A_2 \times \cdots \times A_k \in B(S^k)$,

$$P_\mu(A_1 \times A_2 \times \cdots \times A_k) = \int_{A_1} d\mu(x_1) \int_{A_2} \pi(x_1, dx_2) \cdots \int_{A_k} \pi(x_{k-1}, dx_k).$$

The coordinate process $\{X_i: i \geq 1\}$ on $\Omega = S^\infty$ is a Markov chain and the empirical law L_n can be written

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

The Polish space X will be the set of probability measures on S with the weak topology and the vector space E containing X will be the set of finite signed measures on S with the topology generated by the bounded continuous functions on S acting as linear functionals on E. We are concerned with the sequence of laws on X given by $P_n = P_\mu \circ L_n^{-1}$. Assume that π is a Feller transition probability and define the operator T_ξ on the Banach space $\mathbf{C}(S)$ of continuous, bounded functions on S by

$$T_\xi(g)(x) = \int_S \exp(\langle \delta_y, \xi \rangle) g(y) \pi(x, dy). \tag{4.1}$$

Define the measure μ_ξ on S by

$$\mu_\xi(A) = \int_A \exp(\langle \delta_y, \xi \rangle) \mu(dy). \tag{4.2}$$

Then we have the following representation for ψ.

THEOREM 4.1. *Suppose that $\{P_\mu \circ L_n^{-1}\}$ has a convex large deviation rate function I. Then for any $\xi \in E'$,*

$$\psi(\xi) = \lim \frac{1}{n} \log \mu_\xi [T_\xi]^n(\mathbf{1})$$

and $I = \psi^*$.

PROOF. Observe that if $M \geq \|\xi\| = \sup_{y \in S} |\xi(y)|$, then

$$\begin{aligned} \phi(\langle \cdot, \xi \rangle \wedge M) &= \lim \frac{1}{n} \log \int_X \exp[n(\langle p, \xi \rangle \wedge M)] P_\mu \circ L_n^{-1}(dp) \\ &= \lim \frac{1}{n} \log \int_X \exp[n \langle p, \xi \rangle] P_\mu \circ L_n^{-1}(dp) \end{aligned}$$

since the measure $P_\mu \circ L_n^{-1}$ is concentrated on the probability measures and $\langle p, \xi \rangle \leq \|\xi\|$ for every probability measure $p \in E$. Therefore

$$\begin{aligned} \psi(\xi) &= \lim \frac{1}{n} \log \int_X \exp[n \langle p, \xi \rangle] P_\mu \circ L_n^{-1}(dp) \\ &= \lim \frac{1}{n} \log \int_{S^\infty} \exp[n \langle L_n, \xi \rangle] P_\mu(d\omega) \\ &= \lim \frac{1}{n} \log \int_{S^\infty} \exp\left[\sum_{i=1}^n \xi(y_i)\right] P_\mu(d\omega). \end{aligned}$$

Now, de Acosta [(1985), page 557] showed that

$$\int_{S^\infty} \exp\left[\sum_{i=1}^n \xi(y_i)\right] P_\mu(d\omega) = \mu_\xi [T_\xi]^{n-1}(\mathbf{1}),$$

and therefore ψ takes the form stated and it follows from Theorem 3.1 that $I = \psi^*$. □

The limit $\psi(\xi)$ can be related to two familiar quantities in the theory of nonnegative operators. If $r(\xi)$ denotes the spectral radius of the operator T_ξ, it is clear that

$$\lim \frac{1}{n} \log \mu_\xi [T_\xi]^n (\mathbf{1}) \le \log r(\xi).$$

On the other hand, if π is λ-irreducible and $k(\xi)$ denotes the convergence parameter of the operator T_ξ [see Nummelin (1984)], the following lower bound is always valid.

THEOREM 4.2. *Suppose π is λ-irreducible and let T_ξ and μ_ξ be defined as in* (4.1) *and* (4.2). *Then*

$$\lim \frac{1}{n} \log \mu_\xi [T_\xi]^n (\mathbf{1}) \ge \log \left[\frac{1}{k(\xi)} \right].$$

PROOF. We first produce a small measure [see Nummelin (1984), page 14] inferior to $\mu_\xi T_\xi^n$ for some $n \ge 1$. Let ν be any small measure for the kernel T_ξ. Now μ_ξ may be singular with respect to ν, but $\mu_\xi T_\xi^m$ cannot be singular with respect to ν for every $m \ge 1$. Indeed, suppose on the contrary that every measure $\mu_\xi T_\xi^m$, $m \ge 1$, is singular with respect to ν. Let A_m and B_m be a disjoint decomposition of S such that

$$\nu(A_m) = 0, \qquad \mu_\xi T_\xi^m (B_m) = 0.$$

Let $A = \bigcup_m A_m$ and let $B = \bigcap_m B_m$. Then $A \cup B = S$, $A \cap B = \varnothing$ and

$$\nu(A) = 0, \qquad \mu_\xi T_\xi^m (B) = 0$$

for every $m \ge 1$. Now since $B = S - A$, it follows that $\nu(B) > 0$. Now by the irreducibility of T_ξ, it follows that

$$S = \bigcup_m \{x \in S : T_\xi^m (x, B) > 0\},$$

and therefore one of these sets, say $\{x \in S\colon T_\xi^{m_0}(x, B) > 0\}$ has positive μ_ξ-measure. Hence $\mu_\xi T_\xi^{m_0}(B) > 0$, which contradicts the assumption that $\mu_\xi T_\xi^{m_0}$ was singular with respect to ν. Hence one of the measures $\{\mu_\xi T_\xi^m\colon m \ge 1\}$, say $\mu_\xi T_\xi^{m_0}$, is not singular with respect to ν.

Let α_a be the (nontrivial) absolutely continuous part in the Lebesgue decomposition of $\mu_\xi T_\xi^{m_0}$ with respect to ν. Define a new measure γ on S by

$$\gamma(A) = \int_A \frac{d\alpha_a}{d\nu} \wedge 1 \, d\nu.$$

Clearly, $\gamma(S) > 0$, $\gamma \le \nu$, $\gamma \le \mu_\xi T_\xi^{m_0}$ and finally γ is small since

$$s \otimes \nu \ge s \otimes \gamma.$$

Finally, it is obvious that

$$\limsup \frac{1}{n} \log \mu_\xi [T_\xi]^n (\mathbf{1}) \ge \limsup \frac{1}{n} \log \gamma [T_\xi]^n (\mathbf{1})$$

and the right-hand side is superior to $\log(1/k(\xi))$ by Proposition 3.4 of Nummelin (1984) since γ is a small measure. □

We thus have trapped a potential convex rate function I in the following way:

$$[\log r]^* \le I \le \left[\log \frac{1}{k}\right]^*. \tag{4.3}$$

In the example below,

$$[\log r]^*(p) < \psi^*(p) < \left[\log \frac{1}{k}\right]^*(p)$$

for certain $p \in X$ and ψ^* is in fact the large deviation rate function. In particular, neither $[\log(1/k)]^*$ nor $[\log r]^*$ is the rate function.

EXAMPLE 4.1. Consider the Markov chain on the state space $S = \{1, 2, 3\}$ with transition matrix

$$\pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}.$$

We identify the probability measures on S with the probability simplex in $\mathbf{R}^3$. The Markov chain has maximal irreducibility measure δ_3, and since

$$\pi^2 \ge \beta \mathbf{1} \otimes \delta_3,$$

for some $\beta > 0$, it follows from Proposition 3.4 of Nummelin (1984) that if $\xi = (\xi_1, \xi_2, \xi_3)$, then

$$\begin{aligned} \log \frac{1}{k_\xi} &= \lim \frac{1}{n} \log \delta_3 T_\xi^n(\mathbf{1}) \\ &= \lim \frac{1}{n} \log [0, 0, 1] \begin{bmatrix} \frac{1}{2}e^{\xi_1} & \frac{1}{2}e^{\xi_2} & 0 \\ 0 & \frac{1}{2}e^{\xi_2} & \frac{1}{2}e^{\xi_3} \\ 0 & 0 & e^{\xi_3} \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \xi_3. \end{aligned}$$

Hence $[\log(1/k)]^*(\delta_3) = 0$ and $[\log(1/k)]^*(p) = \infty$ for $p \ne \delta_3$.

One can quickly see that if $p = (p_1, p_2, p_3)$ belongs to the probability simplex in $\mathbf{R}^3$, then

$$\begin{aligned} [\log r]^*(p) &= \sup_\xi \left\{ \xi_1 p_1 + \xi_2 p_2 + \xi_3 p_3 - \log \max\left\{ \tfrac{1}{2}e^{\xi_1}, \tfrac{1}{2}e^{\xi_2}, e^{\xi_3} \right\} \right\} \\ &= (p_1 + p_2) \log 2. \end{aligned}$$

Now by starting the Markov chain at state $\{2\}$, neither $[\log(1/k)]^*$ nor $[\log r]^*$ is the large deviation rate function for the empirical probability measure L_n. The correct rate function is the convex conjugate of

$$\psi(\xi) = \lim \frac{1}{n} \log[0,1,0] \begin{bmatrix} \frac{1}{2}e^{\xi_1} & \frac{1}{2}e^{\xi_2} & 0 \\ 0 & \frac{1}{2}e^{\xi_2} & \frac{1}{2}e^{\xi_3} \\ 0 & 0 & e^{\xi_3} \end{bmatrix}^n \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \log \max\{\tfrac{1}{2}e^{\xi_2}, e^{\xi_3}\},$$

which for (p_1, p_2, p_3) on the probability simplex in $\mathbf{R}^3$ is

$$\psi^*(p_1, p_2, p_3) = \begin{cases} p_2 \log 2, & \text{if } p_1 = 0, \\ \infty, & \text{if } p_1 > 0. \end{cases}$$

It is not hard to see that if the Markov chain starts at {1}, then $[\log r]^*$ is the large deviation rate function, whereas if it starts at {3}, then $[\log(1/k)]^*$ is the rate function.

In general, one cannot expect that $[\log(1/k)]^*$ would equal $[\log r]^*$, as Example 4.1 shows, since $\log r(\xi)$ is, roughly speaking, $\lim(1/n)\log \mu_\xi[T_\xi]^n(f)$ when μ_ξ and f are large (in an imprecise sense) whereas $\log(1/k(\xi))$ is, roughly speaking, the same limit when μ_ξ and f are small [in the precise meaning of Nummelin (1984)]. A notion of a large measure μ could be the following: For some $k \geq 0$ and some $\beta > 0$,

$$\mu \sum_{i=1}^{k} \pi^i \geq \beta \pi(x, \cdot) \tag{4.4}$$

for all $x \in S$. Assuming (4.4), if $\{P_\mu \circ L_n^{-1}\}$ has convex large deviation rate function I, then $\psi(\xi) = \log r(\xi)$ and $I = [\log r]^*$. The proof is straightforward and will be omitted.

Let us see how condition (4.4) applies to Example 4.1. The measures satisfying condition (4.4) are those with some mass at {1}, which can be seen by setting $x = 1$ in (4.4). But the measures for which the conclusion that $I = [\log r]^*$ holds are those with some mass at {1}. Therefore the condition is precise when applied to Example 4.1.

Recall that de Acosta [(1990), Theorem 6] showed that if (i) π is irreducible and if (ii) for every π-closed set C and every $x \in C^c$, there exists $k \geq 1$ such that $\pi^k(x, C^c) = 0$, then $[\log(1/k)]^* = [\log r]^*$. Although the Markov chain in Example 4.1 is δ_3-irreducible, the second of the two conditions of de Acosta is not satisfied and $[\log(1/k)]^* > [\log r]^*$. It appears that even in simple cases $[\log(1/k)]^*$ may be strictly greater than $[\log r]^*$ and that a large deviation rate function can often be strictly between $[\log r]^*$ and $[\log(1/k)]^*$.

A discussion of the relationship between r and $1/k$ can be found in de Acosta (1988). De Acosta [(1988), Remark 1, page 952] noted that the existence of an irreducibility measure β with respect to which each transition probability $\pi(x, \cdot)$ is absolutely continuous implies the two conditions of the preceding paragraph. The existence of such a measure β is Hypothesis H of Donsker and Varadhan [(1976), page 410], under which the authors proved that the I-function

$$I(q) = \sup_u \int \log\left(\frac{u}{\pi u}\right) dq,$$

where u ranges over the bounded continuous functions bounded above zero, serves as a lower rate function for L_n. De Acosta [(1985), page 562] showed that in fact the I-function above is equal to $[\log r]^*$.

Consider now an elaboration of Example 4.1 in which for every starting point of the Markov chain the function $[\log r]^*$ is not the large deviation rate, but there is a starting point for which the chain has large deviation rate $[\log(1/k)]^*$.

EXAMPLE 4.2. Let the countable state space S for the Markov chain be $\{1, 2, 3, \dots\}$ and let the transition probabilities be given by

$$\pi_{1,1} = 1,$$
$$\pi_{i,i-1} = \tfrac{1}{2}, \qquad i \geq 2,$$
$$\pi_{i,i} = \tfrac{1}{2}, \qquad i \geq 2.$$

The transition kernel is δ_1-irreducible and if $\xi = (\xi_1, \xi_2, \dots)$ is a bounded continuous function on S, then

$$\log \frac{1}{k(\xi)} = \lim n^{-1} \log \delta_1 T_\xi^n(1_{\{1\}}) = \xi_1.$$

Hence $[\log(1/k(\xi))]^*(\delta_1) = 0$, but $[\log(1/k(\xi))]^*(p) = \infty$ for $p \neq \delta_1$.

It is easy to see that

$$\log r(\xi) = \log\left[\sup\left\{e^{\xi_1}, \tfrac{1}{2}e^{\xi_2}, \tfrac{1}{2}e^{\xi_3}, \dots\right\}\right]$$

and hence $[\log r]^*(p) = (1 - p_1)\log 2$ [set $\xi = (1 - \log 2, 1, 1, \dots)$ in the expression $[\log r]^*(p) = \sup_\xi\{\langle p, \xi\rangle - \log r(\xi)\}$ to attain this value]. If the chain starts at the point $k > 1$, then it is easily shown that the large deviation rate I_k is

$$I_k(p) = \begin{cases} (p_2 + p_3 + \cdots + p_k)\log 2, & p_i = 0, i > k, \\ \infty, & \text{otherwise,} \end{cases}$$

which is clearly neither $[\log r]^*$ nor $[\log(1/k)]^*$.

To formulate Proposition 4.1 we define a *minimal closed set* to be a π-closed and topologically closed set C_0 with the property that

$$\pi(x, C^c) = 0 \tag{4.5}$$

for every $x \in C_0$ and every π-closed set C. For example, a π-closed and topologically closed set C_0 that is contained in every other π-closed set is minimal. In Examples 4.1 and 4.2, the minimal closed set is a singleton. Under

Hypothesis H of Donsker and Varadhan (1976), the support of the irreducibility measure β is a minimal closed set.

Whereas Example 4.2 shows that even in simple cases the sequence $\{P_z \circ L_n^{-1}\}$ may have a large deviation rate different than $[\log r]^*$ for any starting point $z \in S$, this cannot happen for $[\log(1/k)]^*$ in the presence of a minimal closed set. The next result says that when there exists a minimal closed set, every starting point $z \in C_0$ begins a Markov chain such that the sequence $\{P_z \circ L_n^{-1}\}$ can only have $[\log(1/k)]^*$ as a rate function.

Proposition 4.1. *Let π be λ-irreducible and suppose there is a minimal closed set $C_0 \subset S$. Then the sequence $\{P_\mu \circ L_n^{-1}\}$ with any initial distribution μ such that $\mu(C_0) = 1$ has large deviation rate function $I = [\log(1/k)]^*$ if in fact it has a convex large deviation rate function.*

Proof. Let I be the large deviation rate for the Markov chain with initial distribution μ. Let ν be a probability measure on S. If $\nu(C_0^c) > 0$, then since C_0^c is open, there is a weak neighborhood N_ν of ν with $q(C_0^c) > 0$ for all $q \in N_\nu$. But by Remark 2.1, $I(\nu) = \infty$. On the other hand, de Acosta [(1990), page 417] showed that $[\log(1/k)]^*(\nu) = \infty$ when $\nu(C_0^c) > 0$ and C_0 is π-closed.

Assume now that $\nu(C_0) = 1$ and let ν_0 be the restriction of ν to C_0. Define the operator $T_{\xi,0}$: $\mathbf{C}(C_0) \to \mathbf{C}(C_0)$ by

$$T_{\xi,0}(g)(x) = \int_{C_0} \exp(\langle \delta_y, \xi \rangle) g(y) \pi(x, dy).$$

Let $r_0(\xi)$ denote the spectral radius of $T_{\xi,0}$ and let $k_0(\xi)$ denote the convergence parameter of the kernel $T_{\xi,0}$. (Note that if π is λ-irreducible, then $T_{\xi,0}$ is λ-irreducible and k_0 is defined.)

Observe next that

$$\left[\log \frac{1}{k_0}\right]^*(\nu_0) = \left[\log \frac{1}{k}\right]^*(\nu).$$

To see this, let $\xi \in \mathbf{C}(S)$ and let ξ_0 denote the restriction of ξ to C_0. Since π is λ-irreducible, any π-closed set C_0 has positive λ-measure and hence from Propositions 2.6 and 3.4 of Nummelin (1984), $k(\xi) = k_0(\xi_0)$. Therefore

$$\int_S \xi \, d\nu - \log \frac{1}{k(\xi)} = \int_{C_0} \xi_0 \, d\nu_0 - \log \frac{1}{k_0(\xi_0)} \le \left[\log \frac{1}{k_0}\right]^*(\nu_0)$$

and $[\log(1/k)]^*(\nu) \le [\log(1/k_0)]^*(\nu_0)$. To show the inequality in the other direction, extend continuous functions on C_0 to all of S and use a similar argument.

Now it follows from the result of de Acosta [(1990), Theorem 6] discussed before Example 4.2 that $[\log(1/k_0)]^* = [\log r_0]^*$ for all probability measures on C_0. One looks at the Markov chain restricted to C_0 and notes that any closed set C for this Markov chain is actually a closed set for the original chain on S, and thus $\pi(x, C^c) = 0$ for all $x \in C_0$ by the minimality of C_0. Thus condition (2) of de Acosta is satisfied and hence the equality follows. Then from (4.3) we have that the large deviation rate function I_0 for the chain restricted to C_0 is $[\log(1/k_0)]^*$. But this means that

$$I(\nu) = I_0(\nu_0) = \left[\log \frac{1}{k_0}\right]^*(\nu_0) = \left[\log \frac{1}{k}\right]^*(\nu)$$

on measures ν with $\nu(C_0) = 1$.

Together with the result when $\nu(C_0) < 1$, it follows that $I(\nu) = [\log(1/k)]^*(\nu)$ for all ν. □

Examples such as those presented here indicate that many interesting questions remain in the study of large deviations of L_n for Markov chains which do not satisfy conventional regularity conditions.

Acknowledgment. I would like to thank the referee for many helpful comments and for the reference Bryc (1990).

REFERENCES

AZENCOTT, R. (1980). Grandes déviations et applications. *Ecole d'Eté de Probabilités de Saint-Flour VIII-1978. Lecture Notes in Math.* **774** 1–176. Springer, New York.

BAHADUR, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

BAHADUR, R. R. and ZABELL, S. L. (1979). Large deviations of the sample mean in general vector spaces. *Ann. Probab.* **7** 587–621.

BRYC, W. (1990). Large deviations by the asymptotic value method. In *Diffusion Processes and Related Problems in Analysis* **1** (M. Pinsky, ed.) 447–472. Birkhäuser, Boston.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations. *Ann. Math. Statist.* **23** 493–507.

CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* **736**.

DE ACOSTA, A. (1985). Upper bounds for large deviations of dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **69** 551–564.

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

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

DINWOODIE, I. H. and ZABELL, S. L. (1992). Large deviations for exchangeable random vectors. *Ann. Probab.* **20** 1147–1166.

DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time—III. *Comm. Pure Appl. Math.* **24** 389–461.

EKELAND, I. and TEMAM, R. (1974). *Analyse convexe et problèmes variationnels*. Dunod, Paris.

ELLIS, R. S. (1984). Large deviations for a general class of random vectors. *Ann. Probab.* **12** 1–12.

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

GÄRTNER, J. (1977). On large deviations from the invariant measure. *Theory Probab. Appl.* **22** 24–39.

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

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

OREY, S. (1986). Large deviations in ergodic theory. In *Seminar on Stochastic Processes, 1984* (E. Çinlar, K. L. Chung and R. K. Getoor, eds.) 195–249. Birkhäuser, Boston.

RUELLE, D. (1973). Statistical mechanics on a compact set with Z^ν action satisfying expansiveness and specification. *Trans. Amer. Math. Soc.* **185** 237–251.

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

VARADHAN, S. R. S. (1966). Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** 261–286.

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

DEPARTMENT OF MATHEMATICS
TULANE UNIVERSITY
NEW ORLEANS, LOUISIANA 70118