1

Introduction

We will present the outlook and some of the main contents of the book.

Let P be a Markov kernel and ν a probability measure on a measurable space $(S, \mathcal{S})$, with $\mathcal{S}$ countably generated. Our basic framework will be the canonical Markov chain with transition kernel P and initial distribution ν, given by $(\Omega = S^{\mathbb{N}_0}, \mathcal{S}^{\mathbb{N}_0}, \mathbb{P}_\nu, \{X_j\}_{j\geq 0})$, where $\mathbb{N}_0$ is the set of nonnegative integers, $\{X_j\}_{j\geq 0}$ are the coordinate functions on $S^{\mathbb{N}_0}$, and $\mathbb{P}_\nu$ is the unique probability measure on $(S^{\mathbb{N}_0}, \mathcal{S}^{\mathbb{N}_0})$ such that $\{X_j\}_{j\geq 0}$ is a Markov chain with transition kernel P and initial distribution ν.

Most of this introduction will be devoted to the empirical measure case. Let $\mathcal{P}(S)$ be the space of probability measures on $(S, \mathcal{S})$, and let $B(S)$ be the space of real-valued bounded $\mathcal{S}$-measurable functions on S. For a set $\mathcal{F} \subset B(S)$, the $\sigma(\mathcal{P}(S), \mathcal{F})$ topology on $\mathcal{P}(S)$, or simply the $\mathcal{F}$ topology, is the topology induced on $\mathcal{P}(S)$ by $\mathcal{F}$, that is, by the action $\mu \mapsto \int g\, d\mu$, $g \in \mathcal{F}$: a net $\{\mu_\alpha\} \subset \mathcal{P}(S)$ converges in this topology to $\mu \in \mathcal{P}(S)$ if and only if for all $f \in \mathcal{F}$,

$$\lim_\alpha \int f\, d\mu_\alpha = \int f\, d\mu.$$

Let $L_n \colon \Omega \to \mathcal{P}(S)$, $n \geq 1$, be the nth empirical measure associated with $\{X_j\}$:

$$L_n = n^{-1} \sum_{j=0}^{n-1} \delta_{X_j},$$

where for $x \in S$, δ_x is point mass at x.

Proposition 1.1 *Assume that P is positive Harris recurrent (Section B.3) and let π be its unique invariant probability measure.*

1. *Assume that $\mathcal{F} \subset B(S)$ is separable for the uniform norm on $B(S)$. Then for any $\nu \in \mathcal{P}(S)$, $\{L_n\}$ converges $\sigma(\mathcal{P}(S), \mathcal{F})$ to $\pi, \mathbb{P}_\nu$-a.s..*

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

2. *Let $(S, \mathcal{S})$ be a separable metric space with its Borel σ-algebra, and let $C_b(S)$ be the space of real-valued bounded continuous functions on S. Then for any $\nu \in \mathcal{P}(S)$, $\{L_n\}$ converges $\sigma(\mathcal{P}(S), C_b(S))$ to π, $\mathbb{P}_\nu$-a.s..*

This result is proved in Appendix A.

One of the main objectives of the present work is to study the large deviations associated with this result; in particular, to determine when the probabilities $\{\mathbb{P}_\nu[L_n \notin B]\}$ decay exponentially and at what rate, where $B \subset \mathcal{P}(S)$ is measurable in a suitable sense and π is an interior point of B in a suitable topology.

We will state now the definition of the large deviation principle in the present context. Let V be a vector space $\subset B(S)$. For a set $B \subset \mathcal{P}(S)$, its V-closure is denoted $\mathrm{cl}_V(B)$ and its V-interior $\mathrm{int}_V(B)$. The σ-algebra $\mathcal{B}(\mathcal{P}(S), B(S))$ is the smallest σ-algebra on $\mathcal{P}(S)$ for which each map $\mu \mapsto \int g\,d\mu$, $g \in B(S)$, is measurable. In the rest of this text, $B \subset \mathcal{P}(S)$ being measurable will mean $B \in \mathcal{B}(\mathcal{P}(S), B(S))$. Note that the map $L_n\colon \Omega \to \mathcal{P}(S)$ is measurable.

Definition 1.2 For $\nu \in \mathcal{P}(S)$, $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ *satisfies the large deviation principle in the V topology with rate function J* if

1. $J\colon \mathcal{P}(S) \to \overline{\mathbb{R}^+}$ is V-lower semicontinuous.
2. For every measurable set $B \subset \mathcal{P}(S)$,

$$\overline{\lim_n}\, n^{-1} \log \mathbb{P}_\nu[L_n \in B] \leq -\inf\{J(\mu)\colon \mu \in \mathrm{cl}_V(B)\}, \tag{1.1}$$

$$\underline{\lim_n}\, n^{-1} \log \mathbb{P}_\nu[L_n \in B] \geq -\inf\{J(\mu)\colon \mu \in \mathrm{int}_V(B)\}. \tag{1.2}$$

Inequality (1.1) (and, respectively, (1.2)) will be referred to as the upper (respectively, lower) bound.

We will say that J is V-*tight* if for each $a \geq 0$,

$$L_a \triangleq \{\mu \in \mathcal{P}(S)\colon J(\mu) \leq a\}$$

is V-compact (we have adopted here the terminology of Rassoul-Agha and Seppäläinen, 2015).

A set $M \subset \mathcal{P}(S)$ is a *uniformity set* for the upper (lower) bound in the V topology with rate function J if for every measurable set $B \subset \mathcal{P}(S)$,

$$\overline{\lim_n}\, n^{-1} \log \sup_{\nu\in M} \mathbb{P}_\nu[L_n \in B] \leq -\inf\{J(\mu)\colon \mu \in \mathrm{cl}_V(B)\} \tag{1.3}$$

and, respectively,

$$\underline{\lim_n}\, n^{-1} \log \inf_{\nu\in M} \mathbb{P}_\nu[L_n \in B] \geq -\inf\{J(\mu)\colon \mu \in \mathrm{int}_V(B)\}. \tag{1.4}$$

In the literature on large deviations for empirical measures of Markov chains, the cases of $(S, \mathcal{S})$ and the vector space $V \subset B(S)$ defining the topology on $\mathcal{P}(S)$ that have been considered are:

I. S is a Polish space (it is understood that $\mathcal{S}$ is its Borel σ-algebra) and $V = C_b(S)$. In this case, the $C_b(S)$ topology on $\mathcal{P}(S)$ is usually called the weak topology.

II. $(S, \mathcal{S})$ is a measurable space (with $\mathcal{S}$ countably generated) and $V = B(S)$. In this case, the $B(S)$ topology on $\mathcal{P}(S)$ is usually called the τ topology. This includes in particular the case when S is a countable set, $\mathcal{S}$ is its power set, and $V = B(S)$ is the set of all real-valued bounded functions on S.

In order to avoid repetitions and to capture the common features of cases I and II above that are relevant to the study of large deviations for empirical measures of Markov chains, we will introduce in Chapters 3 and 4 certain conditions on V which cover both cases.

We turn now to the statement of some of the main results. Throughout the book, it will be understood that $(S, \mathcal{S})$, P, ν, $\mathbb{P}_\nu$ are as in the basic framework and $\mathcal{P}(S)$ is endowed with the V topology, where V is a vector subspace of $B(S)$; further assumptions on V will be stated as needed. The assumption that S is Polish will be specified in some results.

For simplicity, in this introduction we will limit ourselves to the statement of large deviations results in case II above, omitting results on uniformity sets. But we must first introduce several functions from $\mathcal{P}(S)$ to $\overline{\mathbb{R}^+}$ which will play the role of rate functions in suitable contexts.

For $g \in B(S)$, we define

$$\phi(g) = \overline{\lim_n}\, n^{-1} \log \sup_{x \in S} \mathbb{E}_x \left(\exp S_n(g)\right),$$

where $S_n(g) = \sum_{j=0}^{n-1} g(X_j)$. It will sometimes be convenient to consider $T_n(g) = \sum_{j=1}^{n} g(X_j)$. Since

$$\mathbb{E}_x \left(\exp S_n(g)\right) = e^{g(x)} \mathbb{E}_x \left(\exp T_{n-1}(g)\right),$$

we have

$$\phi(g) = \overline{\lim_n}\, n^{-1} \log \sup_{x \in S} \mathbb{E}_x \left(\exp T_n(g)\right).$$

The *transform kernel* associated with the Markov kernel P and $g \in B(S)$ is defined to be

$$K_g(x, A) = \int_A e^{g(y)} P(x, dy), \qquad x \in S,\ A \in \mathcal{S}.$$

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

Since for all $x \in S$, $\mathbb{E}_x(\exp T_n(g)) = K_g^n 1(x)$, we have

$$\phi(g) = \overline{\lim_n}\, n^{-1} \log \sup_{x \in S} K_g^n 1(x),$$

and if K_g is regarded as a bounded linear operator on the space $B(S)$ with the supremum norm, then

$$\phi(g) = \log r(K_g),$$

where $r(K_g)$ is the spectral radius of K_g. For, by the spectral radius formula,

$$\begin{aligned}\log r(K_g) &= \lim_n n^{-1} \log \|K_g^n\| \\ &= \lim_n n^{-1} \log \|K_g^n 1\| = \phi(g).\end{aligned}$$

For $\mu \in \mathcal{P}(S)$, we define

$$\phi^*(\mu) = \sup\left\{\int g\, d\mu - \phi(g) \colon g \in B(S)\right\}.$$

Since $\phi(0) = 0$, we have $\phi^*(\mu) \geq 0$. For $\nu \in \mathcal{P}(S)$, $g \in B(S)$, we define

$$\phi_\nu(g) = \overline{\lim_n}\, n^{-1} \log \mathbb{E}_\nu\,(\exp S_n(g));$$

it is easily shown that

$$\phi_\nu(g) = \overline{\lim_n}\, n^{-1} \log \nu K_g^n 1.$$

If $x \in S$ and $\nu = \delta_x$, we write

$$\begin{aligned}\phi_x(g) &= \overline{\lim_n}\, n^{-1} \log \mathbb{E}_x\,(\exp S_n(g)) \\ &= \overline{\lim_n}\, n^{-1} \log K_g^n 1(x).\end{aligned}$$

For $\mu \in \mathcal{P}(S)$, we define

$$\phi_\nu^*(\mu) = \sup\left\{\int g\, d\mu - \phi_\nu(g) \colon g \in B(S)\right\}.$$

Since $\phi_\nu(0) = 0$, we have $\phi_\nu^*(\mu) \geq 0$.

The function $I \colon \mathcal{P}(S) \to \overline{\mathbb{R}^+}$ is defined as follows:

$$I(\mu) = \sup\left\{\int \log\left(\frac{e^g}{Pe^g}\right) d\mu \colon g \in B(S)\right\}, \qquad \mu \in \mathcal{P}(S).$$

Taking $g = 0$ in this expression, we have $I(\mu) \geq 0$. We also define for $\lambda \in \mathcal{P}(S)$, $\mu \in \mathcal{P}(S)$,

$$I_\lambda(\mu) = \begin{cases} I(\mu) & \text{if } \mu \ll \lambda, \\ \infty & \text{otherwise.}\end{cases}$$

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

Then $I_\lambda = I$ if and only if $\operatorname{dom} I \subset \mathcal{P}(S, \lambda)$, where $\mathcal{P}(S, \lambda) = \{\mu \in \mathcal{P}(S) \colon \mu \ll \lambda\}$ and for a function $J \colon \mathcal{P}(S) \to \overline{\mathbb{R}^+}$,

$$\operatorname{dom} J = \{\mu \in \mathcal{P}(S) \colon J(\mu) < \infty\}.$$

In the next paragraph we will introduce certain functions that involve the irreducibility of P, a condition that will play an essential role in this work (see Appendix B). As will be seen in Chapter 5, irreducibility is necessary for a central formulation of the large deviation principle to hold. The most important case of I_λ is the case when P is irreducible and $\lambda = \psi$, a P-maximal irreducibility probability measure (Appendix B). The symbol ψ will have this meaning throughout the text (except for a slight departure from this convention in Appendices B and C).

If P is irreducible, then so is K_g and its convergence parameter $R(K_g)$ exists (Appendix C). We define for $g \in B(S)$

$$\Lambda(g) = -\log R(K_g).$$

If (s, ν) is a small pair (Appendix B) then by (C.1),

$$\Lambda(g) = \overline{\lim_n}\, n^{-1} \log \nu K_g^n s;$$

one can show that

$$\begin{aligned} \Lambda(g) &= \overline{\lim_n}\, n^{-1} \log \mathbb{E}_\nu \left[(\exp S_n(g))\, s(X_{n-1}) \right] \\ &= \overline{\lim_n}\, n^{-1} \log \mathbb{E}_\nu \left[(\exp T_{n-1}(g))\, s(X_{n-1}) \right]. \end{aligned}$$

If $\mu \in \mathcal{P}(S)$, we define

$$\Lambda^*(\mu) = \sup \left\{ \int g\, d\mu - \Lambda(g) \colon g \in B(S) \right\}.$$

Since $\Lambda(0) \leq 0$, we have $\Lambda^*(\mu) \geq 0$.

All functions defined above are convex.

As we will see in Chapter 3, ϕ_ν^* is a rate function for the upper bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the τ topology: under suitable conditions, (1.1) holds with $V = B(S)$ and $J = \phi_\nu^*$. The rate function ϕ_ν^* is a convex-analytic construct which emerges as a result of an argument involving compactness and an exponential Markov inequality.

In Chapter 2 it is proved that if P is irreducible, then Λ^* is a rate function for the lower bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the τ topology: (1.2) holds with $V = B(S)$ and $J = \Lambda^*$. The rate function Λ^* is a convex-analytic construct the emergence of which is more subtle and requires an argument based on the fundamental

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

minorization property of irreducible kernels, as well as subadditivity and identification arguments.

For P irreducible, it is always the case that $\phi_\nu^* \le \Lambda^*$. Moreover, as will be seen in Chapter 5, if $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ satisfies the large deviation principle in the τ topology with rate function J, then $\phi_\nu^* \le J \le \Lambda^*$. A central issue will be to establish conditions ensuring the equality $\phi_\nu^* = \Lambda^*$.

It is very important to relate these rate functions to the functions I and I_ψ, which are of primary significance and analytically more tractable. In this direction, we have

Proposition 1.3 (part of Theorems 4.1 and 4.2)

1. *$\phi^* = I$ and for all $\nu \in \mathcal{P}(S)$, $\phi_\nu^* \ge I$.*
2. *Let P be irreducible. Then $\Lambda^* = I_\psi$ and for all $\nu \in \mathcal{P}(S)$, $\phi_\nu^* \le I_\psi$.*

We will now state some of the large deviation results. Under the sole assumption of irreducibility, for any $\nu \in \mathcal{P}(S)$ the lower bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the τ topology with rate function $\Lambda^* = I_\psi$ always holds:

Proposition 1.4 (part of Theorem 2.10) *Let P be irreducible and $\nu \in \mathcal{P}(S)$. Then for any measurable set $B \subset \mathcal{P}(S)$,*

$$\begin{aligned}\underline{\lim}_n\, n^{-1} \log \mathbb{P}_\nu[L_n \in B] &\ge -\inf\{\Lambda^*(\mu) : \mu \in \operatorname{int}_\tau(B)\}\\ &= -\inf\{I_\psi(\mu) : \mu \in \operatorname{int}_\tau(B)\}.\end{aligned}$$

The following result gives a necessary and sufficient analytic condition for the upper bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the τ topology with τ-tight rate function ϕ_ν^*.

Proposition 1.5 (part of Theorem 3.2) *Let $\nu \in \mathcal{P}(S)$.*

1. *For every measurable set $B \subset \mathcal{P}(S)$ with compact τ-closure,*

$$\overline{\lim}_n\, n^{-1} \log \mathbb{P}_\nu[L_n \in B] \le -\inf\{\phi_\nu^*(\mu) : \mu \in \operatorname{cl}_\tau(B)\}. \tag{1.5}$$

2. *The following conditions are equivalent:*
   (i) *If $0 \le g_k \in B(S)$ and $g_k \downarrow 0$ pointwise, then $\phi_\nu(g_k) \to 0$.*
   (ii) *For every measurable set $B \subset \mathcal{P}(S)$,* (1.5) *holds and ϕ_ν^* is τ-tight.*

In Chapter 6 we will present several analytic conditions equivalent to 2(i) and in Chapter 7 sufficient conditions for 2(i), hence for 2(ii). We state one such sufficient condition. A set $C \in \mathcal{S}$ is $(P-\tau)$-*tight* if there exists $m \in \mathbb{N}$ such that $\{P^m(x, \cdot) : x \in C\}$ is τ-relatively compact. Let $\tau_C = \inf\{n \ge 1 : X_n \in C\}$.

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

Proposition 1.6 (part of Theorem 7.8) *Let $\nu \in \mathcal{P}(S)$. Assume that for every $b > 0$, there exists a $(P - \tau)$-tight set C such that*

$$\sup_{x \in C} \mathbb{E}_x e^{b\tau} < \infty, \quad \text{where } \tau = \tau_C,$$
$$\mathbb{E}_\nu e^{b\tau} < \infty.$$

Then 2(i)*, hence also* 2(ii)*, of Proposition 1.5 holds.*

In order to obtain a large deviation principle for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ from Propositions 1.4 and 1.5, we must reconcile the lower bound rate function I_ψ and the upper bound rate function ϕ^*_ν. Under the assumption that P is irreducible, Proposition 1.7 provides necessary and sufficient conditions for the large deviation principle for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the τ topology with τ-tight rate function $\phi^*_\nu = I_\psi$. In particular, it states that this large deviation principle is equivalent to the assertion: $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ satisfies the upper bound in the τ topology with a τ-tight rate function J such that $\operatorname{dom} J \subset \mathcal{P}(S, \psi)$.

Proposition 1.7 (part of Theorem 8.1) *Let P be irreducible and $\nu \in \mathcal{P}(S)$.*

1. *The conditions* (1.6)–(1.9) *are equivalent:*

$$\text{If } 0 \le g_k \in B(S) \text{ and } g_k \downarrow 0 \text{ pointwise, then } \phi_\nu(g_k) \to 0. \tag{1.6a}$$
$$\text{If } 0 \le g \in B(S) \text{ and } \int g\, d\psi = 0, \text{ then } \phi_\nu(g) = 0. \tag{1.6b}$$
$$\phi_\nu = \Lambda. \tag{1.7a}$$
$$I_\psi \text{ is } \tau\text{-tight.} \tag{1.7b}$$
$$\{\mathbb{P}_\nu[L_n \in \cdot]\} \text{ satisfies the large deviation principle in the } \tau \text{ topology with } \tau\text{-tight rate function } \phi^*_\nu = I_\psi. \tag{1.8}$$
$$\{\mathbb{P}_\nu[L_n \in \cdot]\} \text{ satisfies the upper bound in the } \tau \text{ topology with a } \tau\text{-tight rate function } J \text{ such that } \operatorname{dom} J \subset \mathcal{P}(S, \psi). \tag{1.9}$$

2. *If any, hence all, of conditions* (1.6)–(1.9) *is satisfied, then P has a unique invariant probability measure π, $\pi \equiv \psi$, and therefore $I_\psi = I_\pi$.*

In the context of Proposition 1.7, suppose that $B \subset \mathcal{P}(S)$ is measurable and $\pi \in \operatorname{int}_\tau(B)$. Since $I_\psi(\mu) = 0$ implies that $\mu = \pi$ (to be shown in Lemma 5.5) and I_ψ is τ-tight, we have

$$a = \inf \{ I_\psi(\mu) : \mu \in \operatorname{cl}_\tau(B^c) \} > 0.$$

Therefore the probabilities $\{\mathbb{P}_\nu[L_n \notin B]\}$ indeed decay exponentially at a specified rate:

$$\overline{\lim_n}\, n^{-1} \log \mathbb{P}_\nu[L_n \notin B] \le -a.$$

This form of convergence of $\{L_n\}$ to π is sometimes called *exponential convergence*.

We turn now to the situation when S is countable and P is *matrix irreducible*: for all $x, y \in S$, $\sum_{n=1}^{\infty} P^n(x, y) > 0$. Equivalently, P is irreducible and counting measure is a P-maximal irreducibility measure. In this situation, a significant simplification occurs: $I_\psi = I$ and, by Proposition 1.3, for all $\nu \in \mathcal{P}(S)$, $\phi_\nu^* = I$.

Proposition 1.8 (part of Theorem 9.3 and Remark 9.4) *Let S be countable and assume that P is matrix irreducible.*

1. *The following conditions are equivalent:*

$$\begin{aligned}&\text{For all } b > 0, \text{ there exists } F \text{ finite, } F \subset S, \text{ such that}\\ &\qquad \text{for all } y \in F,\ \mathbb{E}_y e^{b\tau} < \infty, \text{ where } \tau = \tau_F. \qquad (1.10)\\ &\text{If } 0 \le g_k \in B(S) \text{ and } g_k \downarrow 0 \text{ pointwise, then}\\ &\qquad \text{for all } x \in S,\ \phi_x(g_k) \to 0. \qquad (1.11)\\ &\text{For all } x \in S,\ \{\mathbb{P}_x[L_n \in \cdot]\} \text{ satisfies the large deviation principle}\\ &\qquad \text{in the } \tau \text{ topology with } \tau\text{-tight rate function } I. \qquad (1.12)\end{aligned}$$

2. *If any, hence all, of conditions* (1.10)–(1.12) *is satisfied, then P has unique invariant probability measure π and $\pi(x) > 0$ for $x \in S$.*

Finally, we turn to the case of large deviations for vector-valued additive functionals. Let E be a separable Banach space, $f: S \to E$ a measurable function, $S_n(f) = \sum_{j=0}^{n-1} f(X_j)$. Parallel to Proposition 1.1, we have:

Proposition 1.9 *Assume that P and π are as in Proposition 1.1 and that $\int \|f\|\, d\pi < \infty$. Then if $\pi(f) = \int f\, d\pi$, for any $\nu \in \mathcal{P}(S)$,*

$$\lim_n n^{-1} S_n(f) = \pi(f), \quad \mathbb{P}_\nu\text{-a.s.}.$$

Proposition 1.9 is proved in Appendix A.

The objective of Chapter 11 is to study the large deviations associated with this result; in particular, to determine when the probabilities $\{\mathbb{P}_\nu[n^{-1} S_n(f) \notin G]\}$ decay exponentially and at what rate, where $G \subset E$ is open and $\pi(f) \in G$.

To describe the results on vector-valued functionals, we start with the lower bound. Let $F(S)$ be the space of measurable functions $g: S \to \mathbb{R}$ and, as in the case $g \in B(S)$, for $g \in F(S)$ let

$$K_g(x, A) = \int_A e^{g(y)} P(x, dy), \qquad x \in S,\ A \in \mathcal{S}.$$

If P is irreducible, so is K_g; we denote its convergence parameter by $R(K_g)$ (Appendix C).

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

If $f: S \to E$ is a measurable function and $\xi \in E^*$, the dual space of E, we write

$$K_{f,\xi}(x, A) = K_{\langle f,\xi\rangle}(x, A) = \int_A e^{\langle f(y),\xi\rangle} P(x, dy)$$

and define

$$\Lambda_f(\xi) = -\log R(K_{f,\xi}),$$
$$\Lambda_f^*(u) = \sup\{\langle u, \xi\rangle - \Lambda_f(\xi) : \xi \in E^*\}, \qquad u \in E.$$

Since $\Lambda_f(0) \le 0$, we have $\Lambda_f^*(u) \ge 0$.

Under the sole assumption of irreducibility, we have:

Proposition 1.10 (part of Theorem 11.1) *Let P be irreducible and let $f: S \to E$ be measurable. Then for every $\mu \in \mathcal{P}(S)$ and every open set $G \subset E$,*

$$\underline{\lim}_n\, n^{-1} \log \mathbb{P}_\mu \left[n^{-1} S_n(f) \in G\right] \ge -\inf_{u\in G} \Lambda_f^*(u).$$

In Theorem 11.13 the rate function Λ_f^* will be identified in terms of I_ψ.

To state the upper bound result, as in the case $g \in B(S)$ we define for $g \in F(S)$, $\mu \in \mathcal{P}(S)$,

$$\phi_\mu(g) = \overline{\lim_n}\, n^{-1} \log \mathbb{E}_\mu \left(\exp S_n(g)\right),$$

and for $f: S \to E$ measurable, $\xi \in E^*$,

$$\begin{aligned}\phi_{f,\mu}(\xi) &= \phi_\mu\left(\langle f, \xi\rangle\right)\\ &= \overline{\lim_n}\, n^{-1} \log \mathbb{E}_\mu \left(\exp \langle S_n(f), \xi\rangle\right)\end{aligned}$$

and

$$\phi_{f,\mu}^*(u) = \sup\{\langle u, \xi\rangle - \phi_{f,\mu}(\xi) : \xi \in E^*\}, \qquad u \in E.$$

Since $\phi_{f,\mu}(0) = 0$, we have $\phi_{f,\mu}^*(u) \ge 0$.

Proposition 1.11 (part of Theorem 11.16) *Let $f: S \to E$ be measurable, and let $\mu \in \mathcal{P}(S)$. Then:*

1. *For every $\sigma(E, E^*)$-compact set $F \subset E$,*

$$\overline{\lim_n}\, n^{-1} \log \mathbb{P}_\mu \left[n^{-1} S_n(f) \in F\right] \le -\inf_{u\in F} \phi_{f,\mu}^*(u). \tag{1.13}$$

2. *Assume:*

(i) *For some $m \in \mathbb{N}$, $\{P^m(x, \cdot) \circ f^{-1} : x \in S\}$ is tight.*

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

(ii) *For all $r > 0$,*

$$\int \exp\left(r\|f(y)\|\right)\mu(dy) < \infty, \tag{1.14a}$$

$$\sup_{x\in S}\int \exp\left(r\|f(y)\|\right)P(x,dy) < \infty. \tag{1.14b}$$

Then (a) (1.13) *holds for every closed set $F \subset E$ and (b) $\phi^*_{f,\mu}$ is tight.*

The assumptions can be weakened if E is finite dimensional; see Remark 11.18.

We state now a necessary and sufficient condition for the large deviation principle for $\{\mathbb{P}_\mu[n^{-1}S_n(f) \in \cdot]\}$ (for the definition of the large deviation principle relevant here, see, e.g., Rassoul-Agha and Seppäläinen 2015, p. 21).

Proposition 1.12 (part of Theorem 11.21) *Let $\mu \in \mathcal{P}(S)$. Assume:*

1. *P is irreducible.*
2. *P, f, and μ satisfy the assumptions of Proposition 1.11.*

Then the following conditions are equivalent:

$$\phi_{f,\mu} = \Lambda_f.$$

$\{\mathbb{P}_\mu[n^{-1}S_n(f) \in \cdot]\}$ *satisfies the large deviation principle with tight rate function Λ^*_f.*

In the context of Proposition 1.12, and under certain conditions given in Theorem 11.22 which imply that $\Lambda^*_f(u) = 0$ if and only if $u = \pi(f)$, if $G \subset E$ is open and $\pi(f) \in G$, then

$$b = \inf\left\{\Lambda^*_f(u) \colon u \in G^c\right\} > 0.$$

Therefore the probabilities $\{\mathbb{P}_\mu[n^{-1}S_n(f) \notin G]\}$ do decay exponentially at a specified rate:

$$\overline{\lim_n}\, n^{-1}\log \mathbb{P}_\mu\left[n^{-1}S_n(f) \notin G\right] \le -b.$$

In Theorem 11.27 we give a sufficient condition for the large deviation principle for $\{\mathbb{P}_\nu[n^{-1}S_n(f) \in \cdot]\}$: under a certain assumption on P, and if P and f satisfy 2(i) and (1.14b) of Proposition 1.11, the large deviation principle holds for all petite ν (Appendix B) and all $\nu = \delta_x$, where x belongs to a set in $\mathcal{S}$ of full ψ measure.

In Theorem 11.29 we discuss the relationship between the large deviations for empirical measures and the large deviations for additive functionals.

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

1.1 Outline of the Book

We will now give an outline of some of the main contents of each chapter.

Chapter 2 We prove a lower bound for $\{\mathbb{P}_\nu[n^{-1}S_n(f) \in \cdot]\}$, where $f\colon S \to E$ is a bounded measurable function and E is a separable Banach space, and use it to obtain the lower bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the V topology with rate function $(\Lambda \mid V)^*$, hence with rate function $\Lambda^* = I_\psi$. We also present a class of uniformity sets $M \subset \mathcal{P}(S)$ for the lower bound; the class includes all petite sets (Appendix B). We prove that the function $\Lambda\colon B(S) \to \mathbb{R}$ is $\sigma(B(S), \mathcal{P}(S,\psi))$-lower semicontinuous. This result will play a significant role in Chapters 8 and 11.

Chapter 3 We introduce the assumptions V.1–V.3 for a vector space $V \subset B(S)$ and obtain upper bounds for random probability measures. We give a necessary and sufficient analytic condition for the upper bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the V topology with rate function $(\phi_\nu \mid V)^*$. We discuss the connection between the analytic condition and exponential tightness.

Chapter 4 We introduce the assumptions V.1′–V.4 for a vector space $V \subset B(S)$ and prove that $(\phi \mid V)^* = I$ and $\Lambda^* = I_\psi$. We establish conditions under which $(\phi_\nu \mid V)^* = I_\psi$ and $I = I_\psi$. We also study the relationship between $(\Lambda \mid V)^*$ and Λ^* and conditions for the equality $\phi_M^* = I_\lambda$ in certain cases in which P may not be irreducible.

Chapter 5 We assume that $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ satisfies the large deviation principle in the V topology with a rate function J which is not known a priori and derive several consequences, including bounds on J. The existence and uniqueness of invariant measures is discussed. It is proved that if $\{\mathbb{P}_x[L_n \in \cdot]\}$ satisfies the large deviation principle in the τ topology for every $x \in S$ with a common a priori unknown τ-tight rate function J, then P must be irreducible and $J = I$.

Chapter 6 We study in a more abstract setting the analytic condition introduced in Chapter 3 and refine the results found there. We obtain necessary conditions for the uniformity of a set of initial distributions for the upper bound.

Chapter 7 We obtain different sufficient conditions for the upper bound for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ in the V topology.

https://doi.org/10.1017/9781009053129.002 Published online by Cambridge University Press

Chapter 8 We present several formulations for the large deviation principle in the V topology. These include the large deviation principle for $\{\mathbb{P}_\nu[L_n \in \cdot]\}$ for an arbitrary $\nu \in \mathcal{P}(S)$, conditions for a set $M \subset \mathcal{P}(S)$ to be a uniformity set for both the upper and lower bounds, and the case when the large deviation principle holds for $\{\mathbb{P}_x[L_n \in \cdot]\}$ for all $x \in S$.

Chapter 9 We study the case when S is countable and P is matrix irreducible.

Chapter 10 We present several examples which show boundaries of the general results obtained in the previous chapters. In particular, it is shown that even under the assumption of irreducibility, it is possible for the large deviation principle to hold for $\{\mathbb{P}_x[L_n \in \cdot]\}$ for every $x \in S$ with different rate functions for each x.

Chapter 11 We study large deviations for $\{\mathbb{P}_\nu[n^{-1}S_n(f) \in \cdot]\}$, where f is a measurable function on S taking values in a separable Banach space. The identification of the rate function Λ_f^* is discussed, as well as its zero set. We study the relationship between the large deviation principle for empirical measures and the large deviation principle for additive functionals.

Appendix A–Appendix K The appendices have a two-fold purpose. We present some well-known analytical or probabilistic results in a form suitable for application in the main text; in certain cases, we discuss some useful consequences or variants. In particular, Appendices B and C contain definitions and results related to general Markov chains and irreducible kernels. We also prove some auxiliary results for which we have no ready reference. Some of the results proved in the appendices may be new and possibly of independent interest (e.g., Propositions C.3 and I.1).

1.2 Notes

Large deviations for empirical measures (occupation times) of Markov chains were first studied by Donsker and Varadhan (1975) in the case where S is a compact metric space under very strong conditions on P and later in greater generality in Donsker and Varadhan (1976). The basic rate function I was introduced in Donsker and Varadhan (1975). For an entropy representation of I see, e.g., Rassoul-Agha and Seppäläinen (2015, Theorem 13.2). We will comment on the relation of the results in the present book to Donsker and Varadhan (1976) in later chapters; see the notes to Chapters 2 and 7.

Closely related is Gärtner's work, Gärtner (1977).

Other papers that studied large deviations for empirical measures of Markov chains under very strong conditions are Bolthausen (1987), where the τ topology was introduced in this setting, and Ellis (1988). Chapters on the subject may be found in Stroock (1984), Deuschel and Stroock (1989), Dembo and Zeitouni (1998) (largely along the lines of Deuschel and Stroock), Dupuis and Ellis (1997), den Hollander (2000), Rassoul-Agha and Seppäläinen (2015), and Feng and Kurtz (2006).

The present work has its origins in de Acosta (1985, 1988, 1990, 1994a,b) and de Acosta and Ney (1998, 2014). The papers by Ney and Nummelin (1987), Dinwoodie (1993), Dinwoodie and Ney (1995), and Nummelin's book Nummelin (1984) have been highly influential in the development of our outlook. A previous paper related to Ney and Nummelin (1987) is Iscoe et al. (1985). A feature of our approach which is significantly different from the sources cited in the previous paragraphs is the derivation of the lower bound on a general state space under the sole assumption of irreducibility, including the construction of the rate function from the convergence parameters of the irreducible transform kernels. We also focus on initial distributions and uniformity sets and on the existence and uniqueness of invariant measures and their relation to rate functions.

In several chapters we incorporate important contributions of Wu (2000a,b), particularly in connection to upper bounds.

Our presentation of large deviations for vector-valued additive functionals of a Markov chain is based on and extends de Acosta (1985, 1988) and de Acosta and Ney (1998, 2014).

For some results when the assumption of irreducibility is relaxed, see Wu (2000a,b) and Jiang and Wu (2005). For process-level large deviations, a topic not covered here, see Donsker and Varadhan (1983), Deuschel and Stroock (1989), and Wu (2000a,b).