The Annals of Probability
1988, Vol. 16, No. 3, 925–960

SPECIAL INVITED PAPER

LARGE DEVIATIONS FOR VECTOR-VALUED FUNCTIONALS OF A MARKOV CHAIN: LOWER BOUNDS[1]

BY A. DE ACOSTA

Case Western Reserve University

We obtain lower bounds for large deviations of vector-valued functionals of a Markov chain with general state space. The bounds are expressed in terms of the convergence parameter of certain kernels. An application to empirical measures of Markov chains is given.

1. Introduction. Let $\{X_j, j \geq 0\}$ be a Markov chain with state space S and transition probability π. Let E be a topological vector space and $f: S \to E$. Under certain assumptions on the space E, a certain boundedness assumption on f and the sole assumption of irreducibility on π, we obtain lower bounds of the type

$$\liminf_{n \to \infty} \frac{1}{n} \log P_x\left(\left\{\frac{1}{n} \sum_0^{n-1} f(X_j) \in G\right\}\right) \geq -\inf_{u \in G} \lambda(u), \tag{1.1}$$

where G is an open subset of E, $x \in S$, λ is the convex conjugate of

$$\phi(\xi) = -\log R(K_\xi), \qquad \xi \in E^*,$$

and $R(K_\xi)$ is the convergence parameter of the kernel

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

As an application, we derive a basic result of Donsker and Varadhan [5] on lower bounds for large deviations of occupation times of Markov chains taking values in a Polish space, under somewhat weaker assumptions than those in [5]. Although we use at a certain point a technique from [5], in general terms our methods are quite different from those of that paper.

Our point of view is close to that of the interesting very recent work of Ney and Nummelin [11], in which for the first time the convergence parameter of an irreducible kernel is used to construct lower bounds for large deviations of Markov additive processes. In fact, one of our main results, Theorem 5.7, is an infinite-dimensional generalization of the case of Theorem 1 of [11] when the additive component is given by an $\mathbb{R}^d$-valued functional of a Markov chain (see more details in this regard in the remarks following Theorem 5.7). We share with [11] the framework of recent ideas and results from the theory of Markov chains with general state space, such as the convergence parameter of an irreducible

Received June 1987.

[1]This research was partially supported by an NSF grant. This paper is based on a Special Invited Lecture given at the IMS meeting in Madison, Wisconsin in May 1988.

AMS 1980 *subject classifications.* Primary 60F10, 60J05; secondary 60B12.

Key words and phrases. Large deviations, Markov chains, irreducible kernels, convergence parameter, convex analysis.

kernel and its renewal-theoretic characterization and the existence of invariant functions for R-recurrent irreducible kernels, as presented in the very useful recent book of Nummelin [12]. However, these tools are used in a different way in the present paper, in which we emphasize the continuity and approximation properties of irreducible kernels. Moreover, the systematic use of results from infinite-dimensional convex analysis occupies an important position in our work.

We shall describe next the content of each section.

In Section 2 we establish some notation and present some results on irreducible kernels that will be useful later on. In particular, we prove in Theorem 2.1 a general continuity property of the convergence parameter which appears to be of independent interest.

The purpose of Section 3 is to show that if K is an irreducible kernel (satisfying some additional conditions), then a suitable copy $\tilde{K}$ of K can be approximated monotonically from below by a sequence of irreducible quasi-nilpotent kernels (see Definition 2.2). When applied to the kernels π and K_{ξ}, this approximation scheme will play a crucial role in Section 5.

Section 4 is devoted to the proof of certain analytical properties of the convergence parameter of the kernels K_{ξ}, in terms of which the large deviation bounds are defined.

Section 5 contains the main results of the paper, Theorems 5.4, 5.6 and 5.7. In these results lower bounds of the type (1.1) are established; in Theorem 5.4, E is assumed to be a separable Banach space, while in Theorems 5.6 and 5.7, E is assumed to be a vector space with a weak topology satisfying certain additional conditions.

In Section 6 we apply Theorem 5.6 to the case of occupation times of Markov chains, mentioned earlier in this introduction.

Section 7 contains some remarks on the relationship between the lower bounds obtained in this paper and upper bounds obtained previously, in particular in [4].

Appendix A is devoted to proving an ergodic theorem for Banach space-valued functionals of a Markov chain. The proof is simple (given the result in the real-valued case); since we know of no ready reference, we have included it here for completeness.

In Appendix B we prove a continuity result for the Fenchel transform (convex conjugation) which plays an important role in Section 5. This result, though closely related to certain theorems in the literature, appears to be new and may be of independent interest.

2. Irreducible kernels. In general, we adopt the framework and notation of [12]. Throughout the paper, $(S, \mathscr{S})$ denotes a measurable space; it is assumed that the σ-algebra $\mathscr{S}$ is countably generated. We write $g \in \mathscr{S}$ if g is an $\mathscr{S}$-measurable (extended) real-valued function defined on S.

The product of two kernels and the action of a kernel on measurable functions and measures are defined and denoted in the usual way. For example, if K is a kernel on $(S, \mathscr{S})$, $g \in \mathscr{S}$, then Kg is the measurable function

$$Kg(x) = \int K(x, dy)g(y), \qquad x \in S,$$

and $K(gI_{(\cdot)})$ is the kernel

$$K(gI_{(\cdot)})(x, A) = K(gI_A)(x), \qquad x \in S,\ A \in \mathscr{S}.$$

If μ is a measure on $(S, \mathscr{S})$, then $\mu(g) = \int g\,d\mu$; μK is the measure

$$\mu K(A) = \int \mu(dx)K(x, A), \qquad A \in \mathscr{S};$$

in particular, $\mu(gI_{(\cdot)})$ is the measure

$$\mu(gI_{(\cdot)})(A) = \int gI_A\,d\mu.$$

Finally, gK is the kernel $(gK)(x, A) = g(x)K(x, A)$ and if ν is a measure on $(S, \mathscr{S})$, then $(g \otimes \nu)(x, A) = g(x)\nu(A)$.

It will be assumed that K is a nonnegative kernel on $(S, \mathscr{S})$ such that K^n is σ-finite for all $n \geq 1$ (see [12], page 1). In most instances, this property of K will obviously follow from other assumptions. K is *bounded* if $\sup_{x \in S} K(x, S) < \infty$; K is *Markov* (resp., *sub-Markov*) if $K(x, S) = 1$ [resp., $K(x, S) \leq 1$] for all $x \in S$.

Given an irreducible kernel K ([12], Definition 2.2), the set of all irreducibility measures will be denoted $\mathscr{I}(K)$.

The convergence parameter of an irreducible kernel K will be denoted $R(K)$ (see [12], Definition 3.2, Theorem 3.2 and Proposition 3.4).

Our first result gives an important continuity property of the convergence parameter.

THEOREM 2.1. *For each $j \geq 1$ let $S_j \in \mathscr{S}$ and assume $S_j \uparrow S$. Let $\mathscr{S}_j = \{A \in \mathscr{S}\colon A \subset S_j\}$. Let K be an irreducible kernel on $(S, \mathscr{S})$ and for $j \geq 1$, let K_j be an irreducible kernel on $(S_j, \mathscr{S}_j)$.*

Assume that for all $x \in S$, $A \in \mathscr{S}$, $K_j(x, A \cap S_j) \uparrow K(x, A)$. Then

$$R(K_j) \downarrow R(K). \tag{2.1}$$

PROOF. By [12], Definition 2.3 and Theorem 2.1, there exist $m \geq 1$ and a small function $s_1 \in \mathscr{S}_1$ and a measure $\nu_1 \neq 0$ on $(S_1, \mathscr{S}_1)$ such that

$$K_1^m \geq s_1 \otimes \nu_1. \tag{2.2}$$

We define the functions s_j on S_j, s on S by

$$s_j(x) = \begin{cases} s_1(x), & x \in S_1, \\ 0, & x \in S_j \sim S_1, \end{cases}$$

$$s(x) = \begin{cases} s_1(x), & x \in S_1, \\ 0, & x \in S \sim S_1, \end{cases}$$

and the measures ν_j on $(S_j, \mathscr{S}_j)$ [resp., ν on $(S, \mathscr{S})$] by $\nu_j(A) = \nu_1(A \cap S_1)$ [resp., $\nu(A) = \nu_1(A \cap S_1)$]. Then it is easily verified that $s_j \otimes \nu_j$ is an atom for K_j^m (see [12], Definition 4.4) and $s \otimes \nu$ is an atom for K^m. It is possible to choose $m = m_0$ and s_1, ν_1 so that $\nu_1(s_1) > 0$ [see [12], Remark 2.1(ii), page 16]. Then it follows that for each $j \geq 1$, the periods of K_j, $j \geq 1$, and K are not larger than m_0 (see [12], pages 20 and 21).

Next, choose $m = m_1$ to be a prime number larger than m_0 and such that (2.2) holds for suitable choices of s_1 and ν_1. To show that this is possible, choose m, s_1, ν_1 such that (2.2) holds; then for any $p \geq 1$,

$$K^{m+p} \geq (s_1 \otimes \nu_1)K^p = s_1 \otimes (\nu_1 K^p).$$

By irreducibility, $K(x, S) > 0$ for all $x \in S$; hence, $K^p(x, S) > 0$ for all $x \in S$ and $\nu_1 K^p \neq 0$. Now choose p so that $m_1 = m + p$ is prime and larger than m_0.

Thus we may assume that (2.2) holds for a number m that is relatively prime with the period of each of the kernels K_j, $j \geq 1$, and K. It follows from [12], Proposition 2.9, that K_j^m, $j \geq 1$, and K^m are irreducible. We shall prove

$$R(K_j^m) \downarrow R(K^m). \tag{2.3}$$

Since by the proof of Proposition 3.5 in [12], $R(K_j^m) = (R(K_j))^m$ and $R(K^m) = (R(K))^m$, (2.1) follows.

To prove (2.3), we define next the kernels

$$K_j'(x, A) = \begin{cases} K_j(x, A \cap S_j), & x \in S_j, \\ 0, & x \in S \sim S_j,. \end{cases}$$

for $x \in S$, $A \in \mathscr{S}$. Then it is easily verified that for $x \in S$, $A \in \mathscr{S}$, $n \geq 1$,

$$K_j'^n(x, A) = \begin{cases} K_j^n(x, A \cap S_j), & x \in S_j, \\ 0, & x \in S \sim S_j, \end{cases}$$

and

$$(K_j'^m - s \otimes \nu)^n(x, A) = \begin{cases} (K_j^m - s_j \otimes \nu_j)^n(x, A \cap S_j), & x \in S_j, \\ 0, & x \in S \sim S_j. \end{cases}$$

Therefore, for $n \geq 1$,

$$\nu(K_j'^m - s \otimes \nu)^n s = \nu_j(K_j^m - s_j \otimes \nu_j)^n s_j. \tag{2.4}$$

Now the assumption implies $K_j'(x, A) \uparrow K(x, A)$ for all $x \in S$, $A \in \mathscr{S}$, and consequently (see, e.g., [15], page 231), for all $n \geq 1$,

$$\begin{aligned} (K_j'^m - s \otimes \nu)^n &\uparrow (K^m - s \otimes \nu)^n, \\ \nu(K_j'^m - s \otimes \nu)^n s &\uparrow \nu(K^m - s \otimes \nu)^n s. \end{aligned} \tag{2.5}$$

By [12], Proposition 4.7,

$$R(K_j^m) = \sup\{r \geq 0: b_j(r) \leq 1\},$$
$$R(K^m) = \sup\{r \geq 0: b(r) \leq 1\},$$

where

$$b_j(r) = \sum_0^\infty r^{n+1} \nu_j (K_j^m - s_j \otimes \nu_j)^n s_j,$$

$$b(r) = \sum_0^\infty r^{n+1} \nu (K^m - s \otimes \nu)^n s$$

(observe that strict inequality appears in [12]; it is easily seen that the present formulation is equivalent). By (2.4),

$$b_j(r) = \sum_0^\infty r^{n+1}\nu\big(K_j'^m - s \otimes \nu\big)^n s.$$

Since $b_j \le b_{j+1} \le b$ for all $j \ge 1$, it is clear that $R(K_j^m) \downarrow l$ (say) and $l \ge R(K^m)$. In order to prove (2.3) we must show

$$l \le R(K^m). \tag{2.6}$$

Given $\varepsilon > 0$ and setting

$$g_j(n) = (R(K^m) + \varepsilon)^{n+1}\nu\big(K_j'^m - s \otimes \nu\big)^n s,$$

$$g(n) = (R(K^m) + \varepsilon)^{n+1}\nu(K^m - s \otimes \nu)^n s,$$

(2.5) and the monotone convergence theorem imply

$$1 < \sum_0^\infty g(n) = \sum_0^\infty \lim_j g_j(n) = \lim_j \sum_0^\infty g_j(n),$$

and therefore there exists j_0 such that for $j \ge j_0$,

$$1 < b_j(R(K^m) + \varepsilon),$$

which implies $R(K_j^m) < R(K^m) + \varepsilon$. Therefore $l \le R(K^m) + \varepsilon$; since ε is arbitrary, (2.6) follows. □

Definition 2.2 and Lemma 2.3 isolate a condition that will be significant in our work.

DEFINITION 2.2. Let K be an irreducible kernel on $(S, \mathscr{S})$ with an atom $s \otimes \nu$. K is *quasinilpotent* (relative to the atom $s \otimes \nu$) if for some $m \ge 1$,

$$(K - s \otimes \nu)^{m+1} = 0. \tag{2.7}$$

LEMMA 2.3. *Let K be an irreducible kernel on $(S, \mathscr{S})$ with an atom $s \otimes \nu$. Assume that K is quasinilpotent. Then*:

(i) $R(K) > 0$, *K is R-recurrent* ([12], *Definition* 3.2) *and*

$$h = \sum_0^m (R(K))^{n+1}(K - s \otimes \nu)^n s$$

[*where m is as in* (2.7)] *is an $R(K)$-invariant function for K* ([12], *Definition* 5.1). *Moreover, $h > 0$ everywhere.*

(ii) *If K is bounded, then so is h.*

(iii) *Assume that K is bounded and define*

$$Q(x, A) = R(K)(h(x))^{-1}K(hI_A)(x), \qquad x \in S,\ A \in S.$$

Then Q is a Harris-recurrent Markov kernel (*see Appendix* A) *with a unique invariant probability measure.*

PROOF. By [12], Proposition 4.7, $R(K) = \sup\{r \geq 0\colon b(r) \leq 1\}$, where

$$b(r) = \sum_0^\infty r^{n+1}\nu(K - s \otimes \nu)^n s = \sum_0^m r^{n+1}\nu(K - s \otimes \nu)^n s.$$

It follows from the form of b that $R(K) > 0$ and $b(R(K)) = 1$; hence K is R-recurrent ([12], Proposition 4.7). The expression for h follows from Theorem 5.1 of [12] and (2.7). The positivity of h follows from [12], Proposition 5.1. This proves (i).

Assertion (ii) is clear from the form of h.

(iii) By (ii), h is bounded. It follows from [12], Proposition 5.4, that Q is a Harris-recurrent Markov kernel. Let $R = R(K)$, $s_1 = Rsh^{-1}$ and $\nu_1 = \nu(hI_{(\cdot)})$. Then $s_1 \otimes \nu_1$ is an atom for Q and the formula

$$(Q - s_1 \otimes \nu_1)^n(x, A) = R^n(h(x))^{-1}(K - s \otimes \nu)^n(hI_A)(x), \qquad x \in S,\ A \in \mathscr{S},$$

is easily proved by induction. Therefore $(Q - s_1 \otimes \nu_1)^{m+1} = 0$. It follows now from [12], Corollary 5.2, that Q has a finite invariant measure, hence a unique invariant probability measure. □

3. Approximation of certain irreducible kernels by quasinilpotent kernels. The constructions in this section are inspired by the idea of split chain (see [12], Sections 4.3 and 4.4 and the references on page 143) and by the truncation-killing technique in [11]. However, our procedure is different and more analytic in spirit; in particular, we do not use the notion of regeneration time.

The first lemma shows that if a kernel K on $(S, \mathscr{S})$ admits a certain decomposition, then it is possible to construct a new kernel $\tilde{K}$ on $(S \times T, \mathscr{S} \otimes \mathscr{T})$, where $(T, \mathscr{T})$ is a measurable space, which is in a certain sense a copy of K. If the kernel K is irreducible and has an atom (satisfying certain conditions), then a suitably chosen $(T, \mathscr{T})$ will provide enough structure to make possible the approximation of a suitably chosen $\tilde{K}$ by well-behaved kernels on $(S \times T, \mathscr{S} \otimes \mathscr{T})$; this is the content of Lemmas 3.2 and 3.3.

LEMMA 3.1. *Let K, K_0, K_1 be kernels on $(S, \mathscr{S})$. Assume that*

$$K(x, A) = v_0(x)K_0(x, A) + v_1(x)K_1(x, A), \qquad x \in S,\ A \in \mathscr{S},$$

where $v_i \in \mathscr{S}$, $0 \leq v_i \leq 1$ $(i = 0, 1)$ and $v_0 + v_1 = 1$.

Let $(T, \mathscr{T})$ be a measurable space and $\{T_0, T_1\}$ a measurable partition of T. Let M_{ij} $(i, j = 0, 1)$ be Markov kernels defined on $(T, \mathscr{T})$ such that

$$M_{00}(y, T_1) = M_{01}(y, T_0) = M_{10}(y, T_1) = M_{11}(y, T_0) = 0, \qquad y \in T.$$

Define on $(S \times T, \mathscr{S} \otimes \mathscr{T})$ the kernel

$$\begin{aligned}\tilde{K}((x, y), \cdot) = I_{T_0}(y)\{K_0(v_0I_{(\cdot)})(x) \otimes M_{00}(y, \cdot) + K_0(v_1I_{(\cdot)})(x) \otimes M_{01}(y, \cdot)\}\\ + I_{T_1}(y)\{K_1(v_0I_{(\cdot)})(x) \otimes M_{10}(y, \cdot)\\ + K_1(v_1I_{(\cdot)})(x) \otimes M_{11}(y, \cdot)\}.\end{aligned}$$

Then:

(i) *For every measurable* g: $S^{n+1} \to \mathbb{R}^+$, $n \geq 1$, *and for* $i = 0, 1$,

$$\int \tilde{K}((x_0, y_0), d(x_1, y_1)) \int \cdots \int \tilde{K}((x_{n-1}, y_{n-1}), d(x_n, y_n)) I_{T_i}(y_n) g(x_0, \dots, x_n)$$
$$= I_{T_0}(y_0) \int K_0(x_0, dx_1) \int K(x_1, dx_2)$$
$$\times \int \cdots \int K(x_{n-1}, dx_n) v_i(x_n) g(x_0, \dots, x_n)$$
$$+ I_{T_1}(y_0) \int K_1(x_0, dx_1) \int K(x_1, dx_2)$$
$$\times \int \cdots \int K(x_{n-1}, dx_n) v_i(x_n) g(x_0, \dots, x_n).$$

(ii) *For any measure* μ *on* $(S, \mathscr{S})$, *if* μ_0, μ_1 *are probability measures on* $(T, \mathscr{T})$ *such that* $\mu_0(T_1) = \mu_1(T_0) = 0$ *and*

$$\tilde{\mu} = \mu(v_0 I_{(\cdot)}) \otimes \mu_0 + \mu(v_1 I_{(\cdot)}) \otimes \mu_1,$$

then

$$\int \tilde{\mu}(d(x_0, y_0)) \int \tilde{K}((x_0, y_0), d(x_1, y_1))$$
$$\times \int \cdots \int \tilde{K}((x_{n-1}, y_{n-1}), d(x_n, y_n)) g(x_0, \dots, x_n)$$
$$= \int \mu(dx_0) \int K(x_0, dx_1) \int \cdots \int K(x_{n-1}, dx_n) g(x_0, \dots, x_n).$$

(iii) *For every* $n \geq 1$, $i = 0, 1$, $\tilde{\mu} \tilde{K}^n I_{T_i} = \mu K^n v_i$.

PROOF. (i) is proved by a straightforward induction argument, which we omit. To prove (2), we first observe that from (i) we obtain, for $i = 0, 1$,

$$\int \tilde{\mu}(d(x_0, y_0)) \int \tilde{K}((x_0, y_0), d(x_1, y_1))$$
$$\times \int \cdots \int \tilde{K}((x_{n-1}, y_{n-1}), d(x_n, y_n)) I_{T_i}(y_n) g(x_0, \dots, x_n)$$
$$= \int \mu(dx_0) v_0(x_0) \int K_0(x_0, dx_1) \int K(x_1, dx_2)$$
$$\times \int \cdots \int K(x_{n-1}, dx_n) v_i(x_n) g(x_0, \dots, x_n) \tag{3.1}$$
$$+ \int \mu(dx_0) v_1(x_0) \int K_1(x_0, dx_1) \int K(x_1, dx_2)$$
$$\times \int \cdots \int K(x_{n-1}, dx_n) v_i(x_n) g(x_0, \dots, x_n)$$
$$= \int \mu(dx_0) \int K(x_0, dx_1) \int \cdots \int K(x_{n-1}, dx_n) v_i(x_n) g(x_0, \dots, x_n).$$

Adding the equalities (3.1) for $i = 0, 1$, we get (ii).

To prove (iii), take $g \equiv 1$ in (3.1); then

$$\int \tilde{\mu}(d(x_0, y_0)) \int \tilde{K}^n((x_0, y_0), d(x, y)) I_{T_i}(y) = \int \mu(dx_0) \int K^n(x_0, dx) v_i(x),$$

which is the assertion. □

In the next lemma $(\mathbb{N}, \mathscr{N})$ will denote the set of nonnegative integers with the σ-algebra of all subsets.

LEMMA 3.2. *Suppose that the kernel K on $(S, \mathscr{S})$ satisfies* (a) $K \geq s \otimes \nu$ *with $\nu(s) > 0$ and $0 \leq s \leq 1$ and* (b) *if $\nu(A) > 0$, then $K(x, A) > 0$ for all $x \in S$; in particular, K is irreducible and $\nu \in \mathscr{I}(K)$.*

Define on $(S \times \mathbb{N}, \mathscr{S} \otimes \mathscr{N})$ the kernel

$$\tilde{K}((x, j), \cdot) = I_0(j)\{\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0 + \nu(tI_{(\cdot)}) \otimes \delta_1\}$$
$$+ I_{0^c}(j)\{H(\bar{s}I_{(\cdot)})(x) \otimes \delta_0 + H(tI_{(\cdot)})(x) \otimes L(j, \cdot)\},$$

where $\bar{s} = s/2$, $t = 1 - \bar{s}$, $H(x, A) = (t(x))^{-1}[K(x, A) - \bar{s}(x)\nu(A)]$ and $L(j, \{k\}) = \delta_{j+1}(k)$. Then:

(i) *For all $(x, j) \in S \times N$,*

$$\tilde{K}((x, j), \cdot) \geq I_0(j)\{\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0 + \nu(tI_{(\cdot)}) \otimes \delta_1\}.$$

(ii) *If $(\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0)(A) > 0$, then $\tilde{K}((x, j), A) > 0$ for all $(x, j) \in S \times N$; in particular, $\tilde{K}$ is irreducible and $\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0 \in \mathscr{I}(\tilde{K})$.*

(iii) $R(\tilde{K}) = R(K)$.

PROOF. (i) is obvious. To prove (ii), we observe that if $A \in \mathscr{S} \otimes \mathscr{N}$ and $A_0 = \{x \in S\colon (x, 0) \in A\}$, then $\nu(\bar{s}I_{(\cdot)})(A_0) = (\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0)(A)$. Assume now that $(\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0)(A) > 0$. If $j = 0$, then for all $x \in S$, $\tilde{K}((x, j), A) \geq (\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0)(A) > 0$. If $j \in \{0\}^c$ and $\bar{s}(x) = 0$, then

$$\left(H(\bar{s}I_{(\cdot)})(x) \otimes \delta_0\right)(A) = H(\bar{s}I_{A_0})(x) = K(\bar{s}I_{A_0})(x) > 0$$

by assumption (b), since $\nu(\bar{s}I_{A_0}) > 0$. If $j \in \{0\}^c$ and $\bar{s}(x) > 0$, then

$$K(\bar{s}I_{A_0})(x) - \bar{s}(x)\nu(\bar{s}I_{A_0}) \geq \bar{s}(x)\nu(\bar{s}I_{A_0}) > 0$$

by assumption (a) and, therefore,

$$\left(H(\bar{s}I_{(\cdot)})(x) \otimes \delta_0\right)(A) = H(\bar{s}I_{A_0})(x) > 0.$$

(iii) By [12], Proposition 3.4,

$$R(K) = \sup\left\{r \geq 0\colon \sum_0^\infty r^n \nu K^n \bar{s} < \infty\right\},$$

$$R(\tilde{K}) = \sup\left\{r \geq 0\colon \sum_0^\infty r^n \tilde{\nu} \tilde{K}^n \tilde{s} < \infty\right\},$$

where $\tilde{\nu} = \nu(\bar{s}I_{(\cdot)}) \otimes \delta_0 + \nu(tI_{(\cdot)}) \otimes \delta_1$ and $\tilde{s} = I_0$. Now Lemma 3.1(iii) yields the conclusion. □

In the next result we shall use the notation $\mathbb{N}_m = \{0, 1, \dots, m\}$, $J_m = \mathbb{N}_m \sim \{0\}$; $\mathscr{N}_m$ is the σ-algebra of all subsets of $\mathbb{N}_m$.

LEMMA 3.3. *Under the assumptions of Lemma 3.2, define on $(S \times \mathbb{N}_m, \mathscr{S} \otimes \mathscr{N}_m)$ the kernel*

$$\tilde{K}_m((x, j), \cdot) = I_0(j)\{\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0 + \nu(tI_{(\cdot)}) \otimes \delta_1\} + I_{J_m}(j)\{H(\bar{s}I_{(\cdot)})(x) \otimes \delta_0 + H(tI_{(\cdot)})(x) \otimes L_m(j, \cdot)\},$$

where $L_m = L|\mathbb{N}_m \times \mathscr{N}_m$. Then:

(i) *If $(\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0)(A) > 0$, then $\tilde{K}_m((x, j), A) > 0$ for all $(x, j) \in S \times \mathbb{N}_m$; in particular, $\tilde{K}_m$ is irreducible and $\nu(\bar{s}I_{(\cdot)}) \otimes \delta_0 \in \mathscr{I}(\tilde{K}_m)$.*

(ii) *Define $\tilde{s}$ on $S \times \mathbb{N}_m$ and $\tilde{\nu}$ on $\mathscr{S} \otimes \mathscr{N}_m$ as at the end of the proof of Lemma 3.2. Then $\tilde{s} \otimes \tilde{\nu}$ is an atom for $\tilde{K}_m$ and $\tilde{K}_m$ is quasinilpotent relative to $\tilde{s} \otimes \tilde{\nu}$:*

$$(3.2) \qquad (\tilde{K}_m - \tilde{s} \otimes \tilde{\nu})^{m+1} = 0.$$

(iii) *$\tilde{K}_m((x, j), A \cap (S \times \mathbb{N}_m)) \uparrow \tilde{K}((x, j), A)$ for every $(x, j) \in S \times \mathbb{N}$, $A \in \mathscr{S} \otimes \mathscr{N}$ and $R(\tilde{K}_m) \downarrow R(\tilde{K})$.*

PROOF. (i) is proved exactly as statement (b) of Lemma 3.2.

(ii) The first assertion is obvious. To prove (3.2), let

$$F((x, j), \cdot) = I_{J_m}(j)H(\bar{s}I_{(\cdot)})(x) \otimes \delta_0,$$
$$G((x, j), \cdot) = I_{J_m}(j)H(tI_{(\cdot)})(x) \otimes L_m(j, \cdot).$$

Clearly $G^n = (H(tI_{(\cdot)}))^n \otimes (I_{J_m}L_m)^n$ for all $n \geq 1$. It is easy to check that for all $j, k \in \mathbb{N}_m$, $1 \leq n \leq m$,

$$(I_{J_m}L_m)^n(j, \{k\}) = I_{J_{m-n+1}}(j)\delta_{j+n}(k).$$

It follows that $(I_{J_m}L_m)^m = 0$ and therefore $G^m = 0$. Next, since $F(F + G) = 0$, we have for $n \geq 1$,

$$(F + G)^n = G(F + G)^{n-1} = G^2(F + G)^{n-2} = \cdots = G^{n-1}(F + G),$$

and therefore $(F + G)^{m+1} = G^m(F + G) = 0$. This proves (3.2).

(iii) For $x \in S$, $j \in \mathbb{N}$, let $m \geq j$. Then $I_{0^c}(j) = I_{J_m}(j)$ and if $k \in \mathbb{N}_m$, then $L_m(j, \{k\}) = L(j, \{k\})$. Hence for $m \geq j$, $A \in \mathscr{S}$, $k \in \mathbb{N}_m$,

$$\tilde{K}_m((x, j), A \times \{k\}) = \tilde{K}((x, j), A \times \{k\}).$$

The second statement follows now from Theorem 2.1. □

4. Properties of the kernels K_ξ. Let E be a locally convex Hausdorff topological vector space and let p be a norm on E stronger than the topology of

E. Let p^* be the dual norm on E^*, the dual of E, defined by $p^*(\xi) = \sup\{|\langle \xi, x\rangle|: p(x) \le 1\}$. Let $f\colon S \to E$ be a measurable p-bounded function

$$\sup_{x\in S} p(f(x)) = c < \infty. \tag{4.1}$$

Let π be a sub-Markov kernel on $(S, \mathscr{S})$ and define, for $\xi \in E^*$,

$$K_\xi(x, A) = \int_A e^{\langle \xi, f(y)\rangle}\pi(x, dy), \qquad x \in S,\ A \in \mathscr{S}. \tag{4.2}$$

LEMMA 4.1. *Suppose that π is an irreducible sub-Markov kernel. Let f, K_ξ be as before. Then*:

(i) *For each $\xi \in E^*$, K_ξ is irreducible.*

(ii) *For $\xi \in E^*$, define $\phi(\xi) = -\log R(K_\xi)$. Then $\phi(E^*) \subset \mathbb{R}$, $\phi(0) \le 0$ and ϕ is p^*-Lipschitz.*

(iii) *ϕ is convex.*

(iv) *If $\pi \ge s \otimes \nu$, then $K \ge s \otimes \nu_\xi$, where $\nu_\xi = \nu(e^{\langle \xi, f\rangle} I_{(\cdot)})$.*

PROOF. (i) The irreducibility of K_ξ is a consequence of the following statement, which is easily proved by induction: If K is a kernel on $(S, \mathscr{S})$ and $g \in \mathscr{S}$, $g > 0$ everywhere, then for all $n \ge 1$, $A \in \mathscr{S}$,

$$\left\{x\colon \left[K(gI_{(\cdot)})\right]^n(x, A) > 0\right\} = \{x\colon K^n(x, A) > 0\}.$$

(ii) The fact that $0 < R(K_\xi) < \infty$ for all $\xi \in E^*$, so that $\phi(E^*) \subset \mathbb{R}$, follows from (4.1), Theorem 3.2 of [12] and the fact that $R(K) > 0$ if K is bounded, which is easily proved. Next, $\phi(0) = -\log R(K_0) = -\log R(\pi) \le 0$ since $R(\pi) \ge 1$ because π is sub-Markov.

Let $\xi, \eta \in E^*$. Then for $x \in S$, $A \in \mathscr{S}$,

$$K_\xi(x, A) = \int_A e^{\langle \xi - \eta, f(y)\rangle} e^{\langle \eta, f(y)\rangle}\pi(x, dy) \le e^{cp^*(\xi - \eta)}K_\eta(x, A)$$

and, therefore, for all $n \ge 0$,

$$e^{-ncp^*(\xi-\eta)}K_\eta^n \le K_\xi^n \le e^{ncp^*(\xi-\eta)}K_\eta^n. \tag{4.3}$$

It follows from (4.3) that if K_η satisfies the minorization condition $M(m_0, \beta, s, \nu)$ ([12], Section 2.3), then K_ξ satisfies $M(m_0, \beta', s, \nu)$ for a certain $\beta' > 0$.

For $\xi \in E^*$, let $a_\xi(r) = \sum_0^\infty r^n \nu K_\xi^n s$. Then (4.3) implies: For $r \ge 0$,

$$a_\eta(e^{-cp^*(\xi-\eta)}r) \le a_\xi(r) \le a_\eta(e^{cp^*(\xi-\eta)}r). \tag{4.4}$$

Since $R(K_\xi) = \sup\{r \ge 0\colon a_\xi(r) < \infty\}$ by Proposition (3.4) of [12], it follows from (4.4) that

$$e^{-cp^*(\xi-\eta)}R(K_\xi) \le R(K_\eta) \le e^{cp^*(\xi-\eta)}R(K_\xi),$$

which implies

$$|\phi(\xi) - \phi(\eta)| \le cp^*(\xi - \eta).$$

(iii) We must prove: If $\alpha_1 + \alpha_2 = 1$, $\alpha_1 > 0$, $\alpha_2 > 0$ and $\xi = \alpha_1\xi_1 + \alpha_2\xi_2$, $\xi_1, \xi_2 \in E^*$, then

$$R(K_\xi) \ge \left(R(K_{\xi_1})\right)^{\alpha_1}\left(R(K_{\xi_2})\right)^{\alpha_2}. \tag{4.5}$$

We first prove: If $g \in \mathscr{S}$, $g \geq 0$, then for all $n \geq 0$,

$$K_\xi^n g \leq \left(K_{\xi_1}^n g\right)^{\alpha_1}\left(K_{\xi_2}^n g\right)^{\alpha_2}. \tag{4.6}$$

We proceed by induction. For $n = 0$ (4.6) is obvious. Assume that (4.6) is true for n. Then by Hölder's inequality,

$$\begin{aligned} K_\xi^{n+1}g(x) &= \int e^{\langle \xi, f(y)\rangle}\pi(x, dy)K_\xi^n g(y) \\ &\leq \int e^{\langle \alpha_1\xi_1+\alpha_2\xi_2, f(y)\rangle}\left(K_{\xi_1}^n g(y)\right)^{\alpha_1}\left(K_{\xi_2}^n g(y)\right)^{\alpha_2}\pi(x, dy) \\ &\leq \left(\int e^{\langle \xi_1, f(y)\rangle}K_{\xi_1}^n g(y)\pi(x, dy)\right)^{\alpha_1}\left(\int e^{\langle \xi_2, f(y)\rangle}K_{\xi_2}^n g(y)\pi(x, dy)\right)^{\alpha_2} \\ &= \left(K_{\xi_1}^{n+1}g(x)\right)^{\alpha_1}\left(K_{\xi_2}^{n+1}g(x)\right)^{\alpha_2}, \end{aligned}$$

proving (4.6). Let $0 < \beta < 1$. Then, setting $R_i = R(K_{\xi_i})$, $i = 1, 2$, by (4.6) and Hölder's inequality,

$$\begin{aligned} a_\xi(\beta R_1^{\alpha_1}R_2^{\alpha_2}) &\leq \sum_0^\infty (\beta R_1^{\alpha_1}R_2^{\alpha_2})^n \int \nu(dy)\left(K_{\xi_1}^n s(y)\right)^{\alpha_1}\left(K_{\xi_2}^n s(y)\right)^{\alpha_2} \\ &\leq \sum_0^\infty \left[\left(\beta^{1/\alpha_1}R_1\right)^n \int \nu(dy)K_{\xi_1}^n s(y)\right]^{\alpha_1}\left[\left(\beta^{1/\alpha_2}R_2\right)^n \int \nu(dy)K_{\xi_2}^n s(y)\right]^{\alpha_2} \\ &\leq \left[a_{\xi_1}\left(\beta^{1/\alpha_1}R_1\right)\right]^{\alpha_1}\left[a_{\xi_2}\left(\beta^{1/\alpha_2}R_2\right)\right]^{\alpha_2} < \infty. \end{aligned}$$

The last statement follows from Proposition 3.4 of [12]. Hence $\beta R_1^{\alpha_1}R_2^{\alpha_2} \leq R(K_\xi)$ for any $0 < \beta < 1$ and (4.5) follows.

(iv) is obvious. □

LEMMA 4.2. *Suppose that π is a sub-Markov irreducible kernel with atom $s \otimes \nu$. Assume that π is quasinilpotent: For some $m \geq 1$,*

$$(\pi - s \otimes \nu)^{m+1} = 0.$$

Let f and K_ξ be defined as in (4.1) *and* (4.2). *Then*:

(i) *For each $\xi \in E^*$, K_ξ is quasinilpotent: $(K_\xi - s \otimes \nu_\xi)^{m+1} = 0$, where ν_ξ is as in Lemma* 4.1.

(ii) *For each $\xi \in E^*$, let*

$$h_\xi = \sum_0^m \left(R(K_\xi)\right)^{n+1}\left(K_\xi - s \otimes \nu_\xi\right)^n s.$$

Then h_ξ is an $R(K_\xi)$-invariant function for K_ξ, $h_\xi > 0$ everywhere and h_ξ is bounded. Moreover, if

$$Q_\xi(x, A) = R(K_\xi)\left(h_\xi(x)\right)^{-1}K_\xi(h_\xi I_A)(x), \qquad x \in S, A \in \mathscr{S},$$

then Q_ξ is a Harris-recurrent Markov kernel with a unique invariant probability measure γ_ξ.

(iii) ϕ *is Gâteaux differentiable (see, e.g.,* [6], *page* 23) *on* (E^*, p^*) *and its Gâteaux derivative at* $\xi \in E^*$ *is given by*

$$D\phi(\xi)(\eta) = \int \langle \eta, f \rangle \, d\gamma_\xi, \qquad \eta \in E^*. \tag{4.7}$$

(iv) *Assume that for every* $\xi \in E^*$, *there exists* $z_\xi \in E$ *such that* $\langle \eta, z_\xi \rangle = \int \langle \eta, f \rangle \, d\gamma_\xi$ *for all* $\eta \in E^*$. *Then* ϕ *is lower semicontinuous for the* $\sigma(E^*, E)$ *topology. In particular, if* E *is a separable Banach and* p *is the norm on* E, *then* ϕ *is always* w^**-lower semicontinuous.*

PROOF. (i) Since

$$(K_\xi - s \otimes \nu_\xi)(x, A) = \int_A e^{\langle \xi, f(y) \rangle} (\pi - s \otimes \nu)(x, dy),$$

the assertion follows from the argument at the beginning of the proof of Lemma 4.1.

(ii) follows from Lemma 2.3.

(iii) We prove first: For all $\xi, \eta \in E^*$,

$$\sup_{x \in S} |\log h_\xi(x) - \log h_\eta(x)| \le (2m+1) c p^*(\xi - \eta). \tag{4.8}$$

As in the proof of Lemma 4.1, we have

$$e^{-cp^*(\xi-\eta)} R(K_\eta) \le R(K_\xi) \le e^{cp^*(\xi-\eta)} R(K_\eta), \tag{4.9}$$

$$\begin{aligned} e^{-ncp^*(\xi-\eta)} (K_\eta - s \otimes \nu_\eta)^n &\le (K_\xi - s \otimes \nu_\xi)^n \\ &\le e^{ncp^*(\xi-\eta)} (K_\eta - s \otimes \nu_\eta)^n. \end{aligned} \tag{4.10}$$

By (4.9) and (4.10),

$$e^{-(2m+1)cp^*(\xi-\eta)} h_\eta \le h_\xi \le e^{(2m+1)cp^*(\xi-\eta)} h_\eta$$

and (4.8) follows. We observe next that for any bounded $g \in \mathscr{S}$,

$$h_\xi^{-1} K_\xi g = \rho(\xi) Q_\xi (h_\xi^{-1} g),$$

where $\rho(\xi) = (R(K_\xi))^{-1}$ and for $0 < t \le 1$,

$$K_{\xi + t\eta} g = K_\xi (e^{t\langle \eta, f \rangle} g).$$

Now, using the fact that $\rho(\xi) h_\xi = K_\xi h_\xi$,

$$\begin{aligned} \rho(\xi + t\eta) - \rho(\xi) &= h_\xi^{-1} \{ \rho(\xi + t\eta)(h_\xi - h_{\xi+t\eta}) + K_\xi(h_{\xi+t\eta} - h_\xi) \\ &\qquad + K_\xi([e^{t\langle \eta, f \rangle} - 1] h_{\xi + t\eta}) \} \\ &= h_\xi^{-1} \rho(\xi + t\eta)(h_\xi - h_{\xi+t\eta}) + \rho(\xi) Q_\xi (h_\xi^{-1} [h_{\xi+t\eta} - h_\xi]) \\ &\qquad + \rho(\xi) Q_\xi (h_\xi^{-1} [e^{t\langle \eta, f \rangle} - 1] h_{\xi+t\eta}). \end{aligned}$$

Since $\gamma_\xi Q_\xi = \gamma_\xi$, taking into account (4.8) and integrating with respect to γ_ξ we get

$$\rho(\xi + t\eta) - \rho(\xi) = \rho(\xi + t\eta) \int h_\xi^{-1}(h_\xi - h_{\xi+t\eta})\, d\gamma_\xi$$
$$+ \rho(\xi) \int h_\xi^{-1}(h_{\xi+t\eta} - h_\xi)\, d\gamma_\xi$$
$$+ \rho(\xi) \int h_\xi^{-1}[e^{t\langle \eta, f\rangle} - 1] h_{\xi+t\eta}\, d\gamma_\xi$$

and, therefore,

$$\rho(\xi + t\eta) - \rho(\xi) = \left[1 - \int h_\xi^{-1}(h_\xi - h_{\xi+t\eta})\, d\gamma_\xi\right]^{-1}$$
$$\times \rho(\xi) \int (e^{t\langle \eta, f\rangle} - 1) h_\xi^{-1} h_{\xi+t\eta}\, d\gamma_\xi.$$

By (4.8) and the dominated convergence theorem,

$$\lim_{t\to 0} t^{-1}[\rho(\xi + t\eta) - \rho(\xi)] = \rho(\xi) \int \langle \eta, f\rangle\, d\gamma_\xi.$$

Since $|\int\langle \eta, f\rangle\, d\gamma_\xi| \le cp^*(\eta)$, it follows that ρ is Gâteaux differentiable on (E^*, p^*) and its Gâteaux derivative is $D\rho(\xi)(\eta) = \rho(\xi)\int\langle \eta, f\rangle\, d\gamma_\xi$. Therefore ϕ is Gâteaux differentiable on (E^*, p^*) and

$$D\phi(\xi)(\eta) = D(\log\rho)(\xi)(\eta) = (\rho(\xi))^{-1}\rho(\xi) \int \langle \eta, f\rangle\, d\gamma_\xi = \int \langle \eta, f\rangle\, d\gamma_\xi.$$

(iv) Since ϕ is Gâteaux differentiable at ξ and convex, it is subdifferentiable at ξ ([6], page 23). If $\{\xi_\alpha\}_{\alpha\in A}$ is a net converging to ξ for the $\sigma(E^*, E)$ topology, we have

$$\phi(\xi_\alpha) \ge \phi(\xi) + \langle \xi_\alpha - \xi, z_\xi\rangle, \qquad \liminf_{\alpha\in A} \phi(\xi_\alpha) \ge \phi(\xi).$$

This proves the $\sigma(E^*, E)$ lower semicontinuity of ϕ. If E is a separable Banach space, then taking z_ξ to be the Bochner integral $\int f\, d\gamma_\xi$ the assumption in statement (iv) is satisfied. □

Corollary 4.3 complements Lemma 4.1 under an additional assumption. We omit the proof, which is implicit in the proof of Theorem 5.1; one of the ingredients is Lemma 4.2(iv).

COROLLARY 4.3. *Suppose that* π *is a sub-Markov irreducible kernel with atom* $s \otimes \nu$, *such that*:

(i) $\pi \ge s \otimes \nu$ *with* $\nu(s) > 0$, $0 \le s \le 1$.
(ii) *If* $\nu(A) > 0$, *then* $\pi(x, A) > 0$ *for all* $x \in S$.

Let E *be a separable Banach space and let* f *be as in* (4.1), *where* p *is the norm in* E. *Let* ϕ *be as in Lemma* 4.1. *Then* ϕ *is* w^**-lower semicontinuous.*

5. Lower bounds for large deviations. We consider first the case in which the functional f takes values in a separable Banach space. In Lemma 5.1 the lower bounds are obtained under certain assumptions which are later removed. The proof rests on the results of Sections 2–4 and on some arguments from convex analysis.

LEMMA 5.1. *Suppose that π is a sub-Markov irreducible kernel with atom $s \otimes \nu$, such that*:

(i) $\pi \geq s \otimes \nu$, *with* $\nu(s) > 0$ *and* $0 \leq s \leq 1$.
(ii) *If* $\nu(A) > 0$, *then* $\pi(x, A) > 0$ *for all* $x \in S$.

Let E be a separable Banach space and let f and K_ξ be as in (4.1) *and* (4.2), *where p is the norm in E. As in Lemma* 4.1, *let* $\phi(\xi) = -\log R(K_\xi)$, $\xi \in E^*$. *Let λ be the convex conjugate of ϕ*:

$$\lambda(x) = \sup_{\xi \in E^*} [\langle \xi, x\rangle - \phi(\xi)], \qquad x \in E.$$

Then for any subprobability measure μ on $(S, \mathscr{S})$ and any open set G in B,

$$\liminf_{n\to\infty} \frac{1}{n} \log \int \mu(dx_0) \int \pi(x_0, dx_1) \int \cdots \int \pi(x_{n-2}, dx_{n-1}) I_G\left(\frac{1}{n} \sum_0^{n-1} f(x_j)\right)$$
$$\geq -\Lambda(G),$$

where $\Lambda(G) = \inf_{u \in G} \lambda(u)$.

PROOF. Let $\tilde{\pi}$ be the kernel on $(S \times \mathbb{N}, \mathscr{S} \otimes \mathscr{N})$ constructed in Lemma 3.2 for $K = \pi$ and the atom $s \otimes \nu$. By Lemmas 3.1 and 3.2,

$$\begin{aligned}(\mathrm{I}) &= \int \mu(dx_0) \int \pi(x_0, dx_1) \int \cdots \int \pi(x_{n-2}, dx_{n-1}) I_G\left(\frac{1}{n} \sum_0^{n-1} f(x_j)\right) \\ &= \int \tilde{\mu}(dz_0) \int \tilde{\pi}(z_0, dz_1) \int \cdots \int \tilde{\pi}(z_{n-2}, dz_{n-1}) I_G\left(\frac{1}{n} \sum_0^{n-1} \tilde{f}(z_j)\right) = (\mathrm{II}),\end{aligned}$$

where $\tilde{\mu} = \mu(\bar{s} I_{(\cdot)}) \otimes \delta_0 + \mu(t I_{(\cdot)}) \otimes \delta_1$, $z_j = (x_j, y_j) \in S \times \mathbb{N}$ and $\tilde{f}(z_j) = f(x_j)$. Let $\tilde{K}_\xi$ be the kernel on $(S \times \mathbb{N}, \mathscr{S} \otimes \mathscr{N})$ constructed in Lemma 3.2 for $K = K_\xi$ and the atom $s \otimes \nu_\xi$. Then it is easily checked that

$$\tilde{K}_\xi(z, A) = \int_A e^{\langle \xi, \tilde{f}(w)\rangle} \tilde{\pi}(z, dw), \qquad z \in S \times \mathbb{N},\ A \in \mathscr{S} \otimes \mathscr{N}.$$

By Lemma 3.2(iii),

$$R(\tilde{K}_\xi) = R(K_\xi). \tag{5.1}$$

For $m \geq 1$, let $\tilde{\pi}_m$ (resp. $\tilde{K}_{\xi, m}$) be the kernel on $(S \times \mathbb{N}_m, \mathscr{S} \otimes \mathscr{N}_m)$ constructed in Lemma 3.3 for $K = \pi$ (resp., $K = K_\xi$). Then

$$\tilde{K}_{\xi, m}(z, A) = \int_A e^{\langle \xi, \tilde{f}(w)\rangle} \tilde{\pi}_m(z, dw), \qquad z \in S \times \mathbb{N}_m,\ A \in \mathscr{S} \otimes \mathscr{N}_m.$$

By Lemmas 3.3 and 4.2, if we define on $S \times \mathbb{N}_m$,

$$h_{\xi,m} = \sum_0^m \left(R\left(\tilde{K}_{\xi,m}\right)\right)^{n+1} \left(\tilde{K}_{\xi,m} - \tilde{s} \otimes \tilde{\nu}_\xi\right)^n \tilde{s},$$

where $\tilde{\nu}_\xi = \nu_\xi(\bar{s}I_{(\cdot)}) \otimes \delta_0 + \nu_\xi(tI_{(\cdot)}) \otimes \delta_1$, and on $(S \times \mathbb{N}_m, \mathscr{S} \otimes \mathscr{N}_m)$,

$$Q_{\xi,m}(z, A) = R\left(\tilde{K}_{\xi,m}\right)\left(h_{\xi,m}(z)\right)^{-1} \tilde{K}_{\xi,m}\left(h_{\xi,m} I_A\right)(z),$$

then $Q_{\xi,m}$ is a Harris-recurrent Markov kernel with a unique invariant probability measure $\gamma_{\xi,m}$. For fixed $\xi \in E^*$, set $\rho = (R(\tilde{K}_{\xi,m}))^{-1}$, $h = h_{\xi,m}$, $Q = Q_{\xi,m}$. Then one may write

$$\tilde{\pi}_m(z, A) = \rho h(z) \int_A e^{-\langle \xi, \tilde{f}(w) \rangle} \left(h(z)\right)^{-1} Q(z, dw). \tag{5.2}$$

By (5.2), we have

$$\begin{aligned}
(\mathrm{II}) &\geq \int \tilde{\mu}(dz_0) \int \tilde{\pi}_m(z_0, dz_1) \int \cdots \int \tilde{\pi}_m(z_{n-2}, dz_{n-1}) I_G\left(\frac{1}{n} \sum_0^{n-1} \tilde{f}(z_j)\right) \\
&= \rho^{n-1} \int \tilde{\mu}(dz_0) h(z_0) \int Q(z_0, dz_1) \int \cdots \int Q(z_{n-2}, dz_{n-1}) \\
&\quad \times \exp\left(-\left\langle \xi, \sum_0^{n-1} \tilde{f}(z_j)\right\rangle\right) \left(h(z_{n-1})\right)^{-1} I_G\left(\frac{1}{n} \sum_0^{n-1} \tilde{f}(z_j)\right) \\
&= (\mathrm{III}).
\end{aligned}$$

Let $a = \inf_z e^{\langle \xi, \tilde{f}(z) \rangle}$, $b = \sup_z h(z)$, $\bar{\mu} = (\tilde{\mu}(h))^{-1} \tilde{\mu}(hI_{(\cdot)})$ and set $C = ab^{-1}\tilde{\mu}(h)$. Then

$$\begin{aligned}
(\mathrm{III}) &\geq C\rho^{n-1} \int \bar{\mu}(dz_0) \int Q(z_0, dz_1) \int \cdots \int Q(z_{n-2}, dz_{n-1}) \\
&\quad \times \exp\left(-\left\langle \xi, \sum_0^{n-1} \tilde{f}(z_j)\right\rangle\right) I_G\left(\frac{1}{n} \sum_0^{n-1} \tilde{f}(z_j)\right) \\
&= (\mathrm{IV}).
\end{aligned}$$

For $m \geq 1$ let $\phi_m(\xi) = -\log R(\tilde{K}_{\xi,m})$, $\xi \in E^*$, and let λ_m be its convex conjugate, defined on E. Let $u \in G$ and assume $\lambda_m(u) < \infty$. Since λ_m is a convex, proper, lower semicontinuous function, by a theorem of Brondsted and Rockafellar [3] (see also [2], Theorem 3, page 262), given $\varepsilon > 0$, there exists $v \in G$ such that $|\lambda_m(v) - \lambda_m(u)| < \varepsilon$ and $\partial\lambda_m(v)$ is not empty, where ∂F is the subdifferential of a function F (see [6], Chapter 1, Section 5). Let $\xi \in \partial\lambda_m(v)$; since ϕ_m is w^*-lower semicontinuous by Lemma 4.2(iv), the duality theorem for conjugate functions ([6], Chapter 1, Propositions 4.1 and 3.1) implies that λ_m^*, the convex conjugate of λ_m, coincides with ϕ_m and, therefore, by [6], Corollary 5.2, page 22, we have $v \in \partial\phi_m(\xi)$. By Lemma 4.2(iii), (iv) and [6], Proposition 5.3, page 23, it follows that

$$v = \int \tilde{f} \, d\gamma_{\xi,m}. \tag{5.3}$$

Let $Q_{\bar{\mu}}$ be the Markovian probability measure on $(S \times \mathbb{N}_m)^{\mathbb{N}}$ determined by the initial distribution $\bar{\mu}$ and the Markov kernel $Q = Q_{\xi, m}$, where ξ is chosen as previously indicated. Let $\{Z_j, j \geq 0\}$ be the coordinate functions on $(S \times \mathbb{N}_m)^{\mathbb{N}}$. Let $V = \{v' \in E: |\langle \xi, v' - v \rangle| < \varepsilon\}$. Then

$$\begin{aligned}(\mathrm{IV}) &\geq C\rho^{n-1} \int \bar{\mu}(dz_0) \int Q(z_0, dz_1) \int \cdots \int Q(z_{n-2}, dz_{n-1}) \\ &\quad \times \exp\left(-\left\langle \xi, \sum_0^{n-1} \tilde{f}(z_j) \right\rangle \right) I_{G \cap V}\left(\frac{1}{n} \sum_0^{n-1} \tilde{f}(z_j) \right) \\ &\geq C\rho^{-1} \exp(n[\log \rho - (\langle \xi, v \rangle + \varepsilon)]) Q_{\bar{\mu}}\left(\left\{ \frac{1}{n} \sum_0^{n-1} \tilde{f}(Z_j) \in G \cap V \right\} \right) \\ &= (\mathrm{V}).\end{aligned}$$

Since $G \cap V$ is an open set containing v, by Theorem A.3 and (5.3) we have

$$\lim_n Q_{\bar{\mu}}\left(\left\{ \frac{1}{n} \sum_{j=0}^{n-1} \tilde{f}(Z_j) \in G \cap V \right\} \right) = 1. \tag{5.4}$$

By [6], Proposition 5.1, page 21,

$$\begin{aligned}\log \rho - \langle \xi, v \rangle &= \phi_m(\xi) - \langle \xi, v \rangle \\ &= \lambda_m(v) \geq \lambda_m(u) - \varepsilon.\end{aligned} \tag{5.5}$$

Now (5.4) and (5.5) imply

$$\liminf_{n \to \infty} \frac{1}{n} \log(\mathrm{I}) \geq \liminf_{n \to \infty} \frac{1}{n} \log(\mathrm{V}) \geq -\lambda_m(u) - 2\varepsilon$$

and since ε is arbitrary, we may conclude: For each $u \in G$ and each $m \geq 1$,

$$\liminf_{n \to \infty} \frac{1}{n} \log(\mathrm{I}) \geq -\lambda_m(u). \tag{5.6}$$

By (5.1) and Lemma 3.3, for every $\xi \in E^*$,

$$\phi_m(\xi) = -\log R\left(\tilde{K}_{\xi, m}\right) \uparrow -\log R\left(\tilde{K}_{\xi}\right) = \phi(\xi). \tag{5.7}$$

The assumptions of Theorem B.3 are satisfied: By Lemmas 4.1 and 4.2, ϕ_m is convex, proper (in fact, real-valued) and w^*-lower semicontinuous. From the Lipschitz property of ϕ_1 it follows that

$$\phi_1(\xi) \geq \phi_1(0) - c\|\xi\|, \qquad \xi \in E^*,$$

and $\phi(0) = -\log R(\pi) \leq 0$, so ϕ is proper. By Theorem B.3 and (5.7) there exists a sequence $\{u_m\} \subset E$ such that $u_m \to u$ in E and

$$\limsup_{m \to \infty} \lambda_m(u_m) \leq \lambda(u). \tag{5.8}$$

It follows from (5.6) and (5.8) that for every $u \in G$,

$$\limsup_{n\to\infty} \frac{1}{n}\log(\mathrm{I}) \geq -\lambda(u).$$

This concludes the proof. □

Lemma 5.2 is a slight modification of a technique in [5], page 412. It will make it possible to drop some of the assumptions in Lemma 5.1.

LEMMA 5.2. *Let π be a sub-Markov kernel on $(S, \mathscr{S})$. For $0 < t < 1$, let*

$$\pi_t = (1-t)\pi \sum_{j=0}^{\infty} (t\pi)^j.$$

Let $\{\tau_j\colon j \geq 1\}$ be a sequence of independent random variables, each geometrically distributed with parameter $(1-t)$,

$$P(\{\tau_j = k\}) = (1-t)t^{k-1}, \qquad k = 1,2,\dots.$$

Assume that $c = \sup_{x\in S}\|f(x)\| < \infty$. For G open in E, $\varepsilon > 0$, let $G_\varepsilon = \{x \in E\colon d(x, G^c) > \varepsilon\}$.

Then for every subprobability measure μ on $(S, \mathscr{S})$,

$$\int\mu(dy_0)\int\pi_t(y_0, dy_1)\int\cdots\int\pi_t(y_{n-2}, dy_{n-1})I_{G_\varepsilon}\left(\frac{1}{n}\sum_{i=0}^{n-1}f(y_i)\right)$$

$$\leq \int\mu(dx_0)\int\pi(x_0, dx_1)\int\cdots\int\pi(x_{n-2}, dx_{n-1})I_G\left(\frac{1}{n}\sum_{j=0}^{n-1}f(x_j)\right)$$

$$+P\left(\left\{\frac{1}{n}(\tau_1 + \cdots + \tau_n) \geq 1 + \frac{\varepsilon}{2c}\right\}\right).$$

PROOF. Let $F\colon S^n \to \mathbb{R}^+$, measurable. We have

$$\int\mu(dy_0)\int\pi_t(y_0, dy_1)\int\cdots\int\pi_t(y_{n-2}, dy_{n-1})F(y_0,\dots, y_{n-1})$$

$$= \sum_{j_1,\dots,j_{n-1}=1}^{\infty}\left(\prod_{k=1}^{n-1}(1-t)t^{j_k-1}\right)\int\mu(dy_0)\int\pi^{j_1}(y_0, dy_1)$$

$$\times\int\cdots\int\pi^{j_{n-1}}(y_{n-2}, dy_{n-1})F(y_0,\dots, y_{n-1}) \tag{5.9}$$

$$= \sum_{j_1,\dots,j_{n-1}=1}^{\infty} P(\{\tau_1 = j_1,\dots, \tau_{n-1} = j_{n-1}\})\int\mu(dx_0)$$

$$\times\int\pi(x_0, dx_1)\int\cdots\int\pi(x_{k_{n-1}-1}, dx_{k_{n-1}})F(x_0, x_{k_1},\dots, x_{k_{n-1}}),$$

where $k_l = \sum_{i=1}^{l} j_i$ for $l \geq 1$.

Let $I = \{0, 1, \ldots, n-1\}$, $J = \{0, k_1, \ldots, k_{n-1}\}$. Then

$$\operatorname{card}(I \vartriangle J) \leq 2(k_{n-1} - n + 1). \tag{5.10}$$

For, clearly $J \sim I \subset [n, k_{n-1}] \cap \mathbb{N}$ and therefore $\operatorname{card}(J \sim I) \leq k_{n-1} - n + 1$. Since $\operatorname{card}(I \sim J) = \operatorname{card}(J \sim I)$, (5.10) follows. Next, by (5.10), setting $k_0 = 0$,

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i=0}^{n-1} f(x_{k_i}) - \frac{1}{n} \sum_{i=0}^{n-1} f(x_i) \right\| &= \frac{1}{n} \left\| \sum_{j \in J \sim I} f(x_j) - \sum_{j \in I \sim J} f(x_j) \right\| \\ &\leq \frac{c}{n} \operatorname{card}(I \vartriangle J) \\ &\leq \frac{2c}{n}(k_{n-1} - n + 1). \end{aligned} \tag{5.11}$$

Let $A = \{(j_1, \ldots, j_{n-1}) \colon j_i \geq 1 \text{ and } (2c/n)(\sum_{i=1}^{n-1} j_i - n + 1) \geq \varepsilon\}$. Then for any $(x_0, x_1, \ldots, x_{k_{n-1}})$, $(j_1, \ldots, j_{n-1})$, by (5.11),

$$I_{G_\varepsilon}\left(\frac{1}{n} \sum_{0}^{n-1} f(x_{k_i})\right) \leq I_G\left(\frac{1}{n} \sum_{0}^{n-1} f(x_i)\right) + I_A((j_1, \ldots, j_{n-1})). \tag{5.12}$$

Let $F(y_0, \ldots, y_{n-1}) = I_{G_\varepsilon}((1/n)\sum_0^{n-1} f(y_i))$. By (5.9) and (5.12),

$$\begin{aligned} &\int \mu(dy_0) \int \pi_t(y_0, dy_1) \int \cdots \int \pi_t(y_{n-2}, dy_{n-1}) I_{G_\varepsilon}\left(\frac{1}{n} \sum_{i=0}^{n-1} f(y_i)\right) \\ &\quad \leq \int \mu(dx_0) \int \pi(x_0, dx_1) \int \cdots \int \pi(x_{n-2}, dx_{n-1}) \\ &\qquad \times I_G\left(\frac{1}{n} \sum_{0}^{n-1} f(x_j)\right) \pi^{k_{n-1}-n+1}(x_{n-1}, S) \\ &\qquad + \sum_{j_1, \ldots, j_{n-1}=1}^{\infty} P(\{\tau_1 = j_1, \ldots, \tau_{n-1} = j_{n-1}\}) I_A((j_1, \ldots, j_{n-1})) \\ &\quad \leq \int \mu(dx_0) \int \pi(x_0, dx_1) \int \cdots \int \pi(x_{n-2}, dx_{n-1}) I_G\left(\frac{1}{n} \sum_{0}^{n-1} f(x_j)\right) \\ &\qquad + P\left(\left\{\frac{1}{n}(\tau_1 + \cdots + \tau_n) \geq 1 + \frac{\varepsilon}{2c}\right\}\right). \end{aligned}$$

□

The next lemma is taken from [5], page 413. The proof is elementary.

LEMMA 5.3. *Let $\{\tau_j, j \geq 1\}$ be as in Lemma 5.2. Then for every $\delta > 0$,*

$$\lim_{t \to 0} \limsup_{n \to \infty} \frac{1}{n} \log P\left(\left\{\frac{1}{n}(\tau_1 + \cdots + \tau_n) \geq 1 + \delta\right\}\right) = -\infty.$$

We come now to one of the main results of the paper. In its statement, P_μ will denote the Markovian probability measure on $S^{\mathbb{N}}$ determined by the Markov kernel π and the initial distribution μ, and $\{X_j,\ j \geq 0\}$ will be the coordinate functions on $S^{\mathbb{N}}$.

THEOREM 5.4. *Suppose that π is an irreducible Markov kernel on $(S, \mathscr{S})$. Let E be a separable Banach space; $f\colon S \to E$ a bounded measurable map. For each $\xi \in E^*$, let*

$$K_\xi(x, A) = \int_A e^{\langle \xi, f(y)\rangle}\pi(x, dy), \qquad \phi(\xi) = -\log\bigl(R(K_\xi)\bigr).$$

Let λ be the convex conjugate of ϕ:

$$\lambda(x) = \sup_{\xi \in E^*}\,[\langle \xi, x\rangle - \phi(\xi)], \qquad x \in E.$$

Then:

(i) $\lambda(u) \geq 0$ *for all* $u \in E$.
(ii) *For every probability measure μ on $(S, \mathscr{S})$ and every open set G,*

$$\liminf_{n\to\infty} \frac{1}{n}\log P_\mu\left(\left\{\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \in G\right\}\right) \geq -\Lambda(G),$$

where $\Lambda(G) = \inf_{u \in G}\lambda(u)$.

PROOF. The first assertion follows at once from the fact that $\phi(0) \leq 0$, which is proved as in Lemma 4.1.

Let $u \in G$ and assume $\lambda(u) < \infty$; otherwise there is nothing to prove. Choose and fix $\varepsilon > 0$ so that $u \in G_\varepsilon$. For π_t as in Lemma 5.2, let

$$A_n(t) = \int \mu(dx_0)\int \pi_t(x_0, dx_1)\int \cdots \int \pi_t(x_{n-2}, dx_{n-1}) I_{G_\varepsilon}\left(\frac{1}{n}\sum_0^{n-1} f(x_j)\right),$$

$$B_n(t) = P\left(\left\{\frac{1}{n}(\tau_1 + \cdots + \tau_n) \geq 1 + \frac{\varepsilon}{2c}\right\}\right),$$

where $c = \sup_{x\in S}\|f(x)\|$ and $\{\tau_j,\ j \geq 1\}$ are as in Lemma 5.2. By Lemma 5.2, we have for any $0 < t < 1$,

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n}\log P_\mu\left(\left\{\frac{1}{n}\sum_0^{n-1} f(X_j) \in G\right\}\right)\\
&\quad= \liminf_{n\to\infty} \frac{1}{n}\log \int \mu(dx_0)\int \pi(x_0, dx_1)\\
&\qquad\times \int \cdots \int \pi(x_{n-2}, dx_{n-1}) I_G\left(\frac{1}{n}\sum_0^{n-1} f(x_j)\right)\\
&\quad\geq \liminf_{n\to\infty} \frac{1}{n}\log[A_n(t) - B_n(t)]^+.
\end{aligned} \tag{5.13}$$

By Theorem 2.1 and Remarks 2.1(i), (ii) of [12], the kernel π_t satisfies assumptions (i) and (ii) of Lemma 5.1. Applying this lemma, for any $0 < t < 1$,

$$\liminf_{n\to\infty} \frac{1}{n} \log A_n(t) \geq -\lambda_t(u), \tag{5.14}$$

where $\lambda_t(u)$ is the convex conjugate of $\phi_t(\xi) = -\log R(K_{\xi,t})$ and

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

Since $\pi_t \geq (1-t)\pi$, it follows that $K_{\xi,t} \geq (1-t)K_\xi$ for all $\xi \in E^*$. Therefore,

$$\begin{aligned} R(K_{\xi,t}) &\leq (1-t)^{-1} R(K_\xi), \\ \phi_t(\xi) &\geq \phi(\xi) + \log(1-t), \\ \lambda_t(u) &\leq \lambda(u) - \log(1-t). \end{aligned} \tag{5.15}$$

Given $1 > \delta > 0$, by Lemma 5.3 we may choose $0 < t < 1$ such that $-\log(1-t) < \delta$ and

$$\limsup_{n\to\infty} \frac{1}{n} \log B_n(t) \leq -(\lambda(u) + 2).$$

Let $n_0 = n_0(t)$ be such that $n \geq n_0$ implies $B_n(t) \leq e^{-(\lambda(u)+1)n}$. By (5.14) and (5.15), we may choose $n_1 = n_1(t)$ such that $n \geq n_1$ implies

$$A_n(t) \geq e^{-n(\lambda(u)+\delta)}.$$

Then for $n \geq \max\{n_0, n_1\}$,

$$\begin{aligned} \frac{1}{n}\log[A_n(t) - B_n(t)]^+ &= \frac{1}{n}\log\left[A_n(t)\left(1 - \frac{B_n(t)}{A_n(t)}\right)\right] \\ &\geq \frac{1}{n}\log A_n(t) + \frac{1}{n}\log(1 - e^{-(1-\delta)n}) \end{aligned}$$

and therefore by (5.14)

$$\liminf_{n\to\infty} \frac{1}{n}\log[A_n(t) - B_n(t)]^+ \geq -(\lambda(u) + \delta).$$

By (5.13), it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log P_\mu\left(\left\{\frac{1}{n}\sum_0^{n-1} f(X_j) \in G\right\}\right) \geq -(\lambda(u) + \delta).$$

Since δ is arbitrary, the proof is complete. □

In the next two results the space E is endowed with a weak topology. Theorem 5.7 actually includes Theorem 5.6, but it seems clearer to state them separately.

DEFINITION 5.5. We shall say that the real vector space E and the functional $f: S \to E$ *satisfy condition* (w) if:

(i) E is endowed with the $\sigma(E, F)$ topology, where F is another real vector space in separating duality with E.

(ii) There is a norm p on E, stronger than $\sigma(E, F)$.

(iii) There exists a closed convex subset $E_1 \subset E$, such that the $\sigma(E, F)$ topology on E_1 is Polish and $p|E_1$ is measurable.

(iv) $f(S) \subset E_1$ and f is measurable.

Assumptions (iii) and (iv) provide one possible way of avoiding measurability problems. Notice also that $(E, \sigma(E, F))^* = F$.

The following is an important example of the situation depicted by Definition 5.5. Let S be a Polish space; $E = \mathscr{M}(S)$, the space of finite signed measures on S; $F = C_b(S)$, the space of bounded continuous functions on S. Then the natural bilinear form $\langle g, \mu\rangle = \int g\, d\mu$ defines a separating duality for the pair (E, F). Let $E_1 = \mathscr{M}_1^+(S)$, the space of probability measures on S; then it is well known that the $\sigma(E, F)$ topology on E_1 is Polish. Let $p = \|\cdot\|_v$, the total variation norm. Finally, let $f: S \to E_1$ be defined by $f(x) = \delta_x$. This example will allow us to apply Theorem 5.6 to the case of occupation times in Section 6.

The proof of Theorem 5.6 is close to that of Lemma 5.1 and Theorem 5.4, but somewhat more elementary, since the deeper convex analysis arguments are carried out in a finite-dimensional space.

THEOREM 5.6. *Suppose that π is an irreducible Markov kernel on $(S, \mathscr{S})$. Assume that the vector space E and the functional f: $S \to E$ satisfy condition (w) and that f is p-bounded.*

Define K_ξ, ϕ, λ as in Theorem 5.4. *Then statements* (i) *and* (ii) *of Theorem* 5.4 *hold, where G is now a $\sigma(E, F)$-open subset of E.*

PROOF. We will indicate the necessary modifications in the proofs of Lemma 5.1 and Theorem 5.4.

Statement (i) is proved as in Theorem 5.4.

1. We first retrace the proof of Lemma 5.1. Defining (I)–(IV) and arguing as in Lemma 5.1, we have

$$\text{(I)} \geq \text{(IV)}. \tag{5.16}$$

Now let $T: E \to \mathbb{R}^d$ be a continuous linear map. Let U be an open set in $\mathbb{R}^d$ such that $T^{-1}(U) \subset G$. For $m \geq 1$ let ϕ_m be as in Lemma 5.1 and define

$$\phi_{m,T}(\alpha) = \phi(T'(\alpha)), \qquad \alpha \in (\mathbb{R}^d)^*,$$

where T': $(\mathbb{R}^d)^* \to E^*$ is the transpose of T, that is, $T'(\alpha) = \alpha \circ T$. Let $\lambda_{m,T}$: $\mathbb{R}^d \to \overline{\mathbb{R}}$ be the convex conjugate of $\phi_{m,T}$. Let $u \in U$ and assume $\lambda_{m,T}(u) < \infty$. By [14], Theorem 23.4 (or by [2], Theorem 3, page 262), given $\varepsilon > 0$ there exists $v \in U$ such that $|\lambda_{m,T}(v) - \lambda_{m,T}(u)| < \varepsilon$ and $\partial\lambda_{m,T}$ is not empty. Let $\alpha \in \partial\lambda_{m,T}$; then by [14], Theorem 23.5 (or by [6], Corollary 5.2, page 22) we have

$v \in \partial\phi_{m,T}(\alpha)$. From Lemma 4.2(iii) it easily follows that $\phi_{m,T}$ is Gâteaux differentiable everywhere (hence, by Theorem 25.2 of [14], even Fréchet differentiable) and

$$D\phi_{m,T}(\alpha) = \int (T \circ \tilde{f}\,)\, d\gamma_{T'(\alpha),\, m}.$$

Now by Theorems 25.1 and 25.2 of [14] (or by [6], Proposition 5.3, page 23) we have

$$v = \int (T \circ \tilde{f}\,)\, d_{T'(\alpha),\, m}. \tag{5.17}$$

Let $V = \{v' \in R^d \colon |\langle \alpha, v' - v\rangle| < \varepsilon\}$. Then taking in (IV) $\xi = T'(\alpha)$, we have

$$\begin{aligned}
\text{(IV)} &\geq C\rho^{n-1}\int \bar{\mu}(dz_0)\int Q(z_0, dz_1)\\
&\quad\times \int\cdots\int Q(z_{n-2}, dz_{n-1})\exp\left(-\left\langle \alpha, \sum_0^{n-1} T\big(\tilde{f}(z_j)\big)\right\rangle\right)\\
&\quad\times I_{T^{-1}(U\cap V)}\left(\frac{1}{n}\sum_0^{n-1}\tilde{f}(z_j)\right)\\
&\geq C\rho^{n-1}\exp\big(n[\log\rho - (\langle\alpha, v\rangle + \varepsilon)]\big)Q_{\bar{\mu}}\left(\left\{\frac{1}{n}\sum_{j=0}^{n-1} T\big(\tilde{f}(Z_j)\big) \in U \cap V\right\}\right)\\
&= \text{(V)}.
\end{aligned} \tag{5.18}$$

By [14], Theorem 23.5 (or by [6], Proposition 5.1, page 21),

$$\begin{aligned}
\log\rho - \langle\alpha, v\rangle &= \phi_{m,T}(\alpha) - \langle\alpha, v\rangle\\
&= \lambda_{m,T}(v) \geq \lambda_{m,T}(u) - \varepsilon.
\end{aligned} \tag{5.19}$$

Now by (5.16)–(5.19) and applying Theorem A.3 as in (5.4), we have

$$\liminf_{n\to\infty}\frac{1}{n}\log(\text{I}) \geq \liminf_{n\to\infty}\frac{1}{n}\log(\text{V}) \geq -\lambda_{m,T}(u) - 2\varepsilon.$$

Since ε is arbitrary, we may conclude: For each T: $E \to \mathbb{R}^d$, for any open set U in $\mathbb{R}^d$ such that $T^{-1}(U) \subset G$ and any $u \in U$, for any $m \geq 1$,

$$\liminf_{n\to\infty}\frac{1}{n}\log(\text{I}) \geq -\lambda_{m,T}(u). \tag{5.20}$$

As in Lemma 5.1, if $\phi_T(\alpha) = \phi(T'(\alpha))$ we have for each $\alpha \in (\mathbb{R}^d)^*$,

$$\phi_{m,T}(\alpha) \uparrow \phi_T(\alpha). \tag{5.21}$$

The assumptions of Theorem B.3 are easily seen to be satisfied. Observe that in this case $\phi_{m,T}$ is continuous on $(\mathbb{R}^d)^*$. By Theorem B.3 (we could invoke at this point the easier finite-dimensional case of this result; see references in [9] and [1]) and (5.21), given $u \in U$, there exists a sequence $\{u_m\} \subset \mathbb{R}^d$ such that

$u_m \to u$ and

$$\limsup_{m\to\infty} \lambda_{m,T}(u_m) \le \lambda_T(u), \tag{5.22}$$

where λ_T is the convex conjugate of ϕ_T. It follows now from (5.20) and (5.22) that

$$\liminf_{n\to\infty} \frac{1}{n}\log(\mathrm{I}) \ge -\lambda_T(u). \tag{5.23}$$

Finally, since G is $\sigma(E, F)$-open, given $x \in G$, one can find a continuous linear map T: $E \to \mathbb{R}^d$ (for some $d \in \mathbb{N}$) and an open set $U \subset \mathbb{R}^d$ such that $x \in T^{-1}(U) \subset G$. Then

$$\begin{aligned} \lambda_T(T(x)) &= \sup_{\beta\in(\mathbb{R}^d)^*} [\langle \beta, T(x)\rangle - \phi_T(\beta)] \\ &= \sup_{\beta} [\langle T'(\beta), x\rangle - \phi(T'(\beta)] \\ &\le \sup_{\eta\in E^*} [\langle \eta, x\rangle - \phi(\eta)] = \lambda(x). \end{aligned} \tag{5.24}$$

Putting now $u = T(x)$, we have from (5.23) and (5.24),

$$\liminf_{n\to\infty} \frac{1}{n}\log(\mathrm{I}) \ge -\lambda(x).$$

2. We retrace now the proof of Theorem 5.4. Let G be a $\sigma(E, F)$-open subset of E. For D finite, $D \subset E^*$, let

$$p_D(u) = \sup_{\xi\in D} |\langle \xi, u\rangle|, \qquad G_D = \left\{u \in E\colon \inf_{v\in G^c} p_D(u-v) > 1\right\}.$$

Then it is easily verified that G_D is $\sigma(E, F)$-open and $\bigcup_D G_D = G$. Let $c = \sup_{x\in S} p(f(x))$ and $d = \sup_{\xi\in D} p^*(\xi)$. Then

$$\sup_{x\in S} p_D(f(x)) \le cd.$$

Let $A_n(t)$ [resp., $B_n(t)$] be defined as in Theorem 5.4, but with G_D instead of G_ε (resp., with $1/2cd$ instead of $\varepsilon/2c$). Then by an obvious modification of Lemma 5.2, we have as in Theorem 5.4: For any $0 < t < 1$, any finite $D \subset E^*$,

$$\liminf_{n\to\infty} \frac{1}{n}\log P_\mu\left(\left\{\frac{1}{n}\sum_0^{n-1} f(X_j) \in G\right\}\right) \ge \liminf_{n\to\infty} \frac{1}{n}\log[A_n(t) - B_n(t)]^+.$$

Given $u \in U$ and choosing D so that $u \in G_D$, we have, by part 1 of the present proof,

$$\liminf_{n\to\infty} \frac{1}{n}\log A_n(t) \ge -\lambda_t(u).$$

The rest of the proof is completed as in Theorem 5.4. □

The next result extends Theorem 5.6 to the case when f may be unbounded. We need, however, an assumption ensuring that the kernel π, when truncated to suitably chosen sets of boundedness of f, remains irreducible.

In the context of Definition 5.5, we shall write $B_r = \{x \in E\colon p(x) \le r\}$.

THEOREM 5.7. *Suppose that* π *is an irreducible Markov kernel on* $(S, \mathscr{S})$. *Assume that the vector space E and the functional* $f\colon S \to E$ *satisfy condition* (w). *Assume furthermore: There exists a positive sequence* $\{r_k\}$, $r_k \uparrow \infty$, *such that for each* $k \ge 1$, π_k *is irreducible on* $(S_k, \mathscr{S}_k)$, *where* $S_k = f^{-1}(B_{r_k})$, $\mathscr{S}_k = \{A \in \mathscr{S}\colon A \subset S_k\}$ *and* $\pi_k(x, A) = \pi(x, A)$, $x \in S$, $A \in \mathscr{S}_k$. *Then the conclusions of Theorem* 5.6 *hold.*

PROOF. It is clear that the proof of Theorem 5.6 is still valid for a sub-Markov kernel and an initial subprobability measure. Let G be $\sigma(E, F)$-open and let $T\colon E \to \mathbb{R}^d$ be a continuous linear map, U an open subset of $\mathbb{R}^d$ such that $T^{-1}(U) \subset G$.

Setting $\mu_k(A) = \mu(A)$ for $A \in \mathscr{S}_k$, we have for any $u \in U$:

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n}\log P_\mu\left(\left\{\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \in G\right\}\right)\\
&\quad\ge \lim_{n\to\infty}\frac{1}{n}\log\int \mu_k(dx_0)\int \pi_k(x_0, dx_1)\\
&\qquad\times\int\cdots\int \pi_k(x_{n-2}, dx_{n-1}) I_U\left(\frac{1}{n}\sum_{0}^{n-1} T(f(x_j))\right)\\
&\quad\ge -\lambda_{T,k}(u),
\end{aligned}$$

where $\lambda_{T,k}\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ is the convex conjugate of $\phi_{T,k}(\alpha) = \phi_k(T'(\alpha))$, $\alpha \in (\mathbb{R}^d)^*$, $\phi_k(\xi) = -\log R(K_{\xi,k})$ and $K_{\xi,k}(x, A) = \int_A e^{\langle \xi, f(y)\rangle}\pi_k(x, dy)$. By the irreducibility of π_k and Lemma 4.1(i), $K_{\xi,k}$ is irreducible on $(S_k, \mathscr{S}_k)$. By Theorem 2.1, for all $\alpha \in (\mathbb{R}^d)^*$,

$$\phi_{T,k}(\alpha) \uparrow \phi_T(\alpha),$$

where $\phi_T(\alpha) = \phi(T'(\alpha))$. It is easily verified using Lemma 4.1 that the assumptions of Theorem B.3 are satisfied for $\{\phi_{T,k}, k \ge 1\}$ and ϕ_T. Applying that theorem, it follows that for $u \in U$ there exists a sequence $\{u_k\} \subset \mathbb{R}^d$ such that $u_k \to u$ and

$$\limsup_{k\to\infty} \lambda_{T,k}(u_k) \le \lambda_T(u).$$

Clearly we have now

$$\liminf_{n\to\infty}\frac{1}{n}\log P_\mu\left(\left\{\frac{1}{n}\sum_{j=0}^{n-1} f(X_j) \in G\right\}\right) \ge -\lambda_T(u).$$

The proof is finished as in part 1 of the proof of Theorem 5.6. □

REMARKS. 1. It appears that a result along the lines of Theorem 5.7 in the case when E is a separable Banach space should be true. There are, however, some technical difficulties in the application of Theorem B.3.

2. As mentioned in the Introduction, if we take $E = \mathbb{R}^d$ in Theorem 5.7 we obtain Theorem 1 of [11] in the case of a functional of a Markov chain. Note, however, that we do not need the minorization condition required there. On the other hand, it appears that the assumption on the truncations of π is needed and should be added in [11].

6. Occupation times. In this section we show how a basic result of Donsker and Varadhan ([5], Theorem 3.3) on lower bounds for large deviations of occupation times of a Markov chain taking values in a Polish space may be derived from Theorem 5.6. In fact, we relax to some extent the absolute continuity assumption in [5] (see Remark 1 following Theorem 6.3).

We shall use the setup described in the example following Definition 5.5.

The first lemma gives a useful alternative expression for the I-functional of [5]. We recall its definition: For $\mu \in \mathscr{M}_1^+(S)$,

$$I(\mu) = \sup_{u \in \mathscr{U}} \int \log\left(\frac{u}{\pi u}\right) d\mu,$$

where $\mathscr{U} = \{u \in \mathscr{S}\colon u \text{ is bounded and } \inf_{x \in S} u(x) > 0\}$.

LEMMA 6.1. *Define for* $\mu \in \mathscr{M}_1^+(S)$,

$$I'(\mu) = \sup\left\{\int \log\left(\frac{u}{\pi u}\right) d\mu\colon u \in \mathscr{S}, u > 0 \textit{ everywhere},\right.$$

$$\left.\mu(\{u = \infty\}) = 0, \left(\log\left(\frac{u}{\pi u}\right)\right)^- \in L^1(\mu)\right\}.$$

Then for all $\mu \in \mathscr{M}_1^+(S)$,

$$I'(\mu) = I(\mu).$$

PROOF. If $u \in \mathscr{U}$, then $\log(u/\pi u)$ is bounded. Hence for any $\mu \in \mathscr{M}_1^+(S)$,

$$I(\mu) \le I'(\mu).$$

To prove the opposite inequality, let $\delta_n > 0$, $\delta_n \downarrow 0$, and define

$$\phi_n(t) = \frac{t}{1 + \delta_n t} + \delta_n, \qquad t \in \mathbb{R}^+,$$

$$\phi_n(+\infty) = \delta_n^{-1} + \delta_n.$$

Then it is easily seen that ϕ_n is increasing, concave, $\delta_n \le \phi_n(t) \le \delta_n + \delta_n^{-1}$, $\lim_n \phi_n(t) = t$ for $t \in \overline{\mathbb{R}}^+$. We claim now: If $u \in \mathscr{S}$, $u > 0$ everywhere, then on $\{u < \infty\}$

$$\log\frac{\phi_n \circ u}{\pi(\phi_n \circ u)} \ge -\left[\log\left(\frac{u}{\pi u}\right)\right]^-. \tag{6.1}$$

First we observe that by Jensen's inequality, for every $x \in S$,

$$\pi(\phi_n \circ u)(x) \le \phi_n(\pi u(x)). \tag{6.2}$$

Suppose $\pi u(x) \le u(x) < \infty$. Then by (6.2),

$$\frac{\phi_n(u(x))}{\pi(\phi_n \circ u)(x)} \ge \frac{\phi_n(u(x))}{\phi_n(\pi u(x))} \ge 1,$$

proving (6.1) in this case. If $u(x) < \pi u(x) = \infty$, then (6.1) is trivial. If $u(x) < \pi u(x) < \infty$, then by (6.2)

$$\begin{aligned}\frac{\phi_n(u(x))}{\pi(\phi_n \circ u)(x)} &\ge \frac{\phi_n(u(x))}{\phi_n(\pi u(x))}\\ &= \frac{(1+\delta_n u(x))^{-1}[u(x)+\delta_n(1+\delta_n u(x))]}{(1+\delta_n \pi u(x))^{-1}[\pi u(x)+\delta_n(1+\delta_n \pi u(x))]}\\ &\ge \frac{u(x)(1+\delta_n^2)+\delta_n}{\pi u(x)(1+\delta_n^2)+\delta_n}\\ &\ge \frac{u(x)}{\pi u(x)}.\end{aligned}$$

To justify the last step, observe that if $0 < a < b < \infty$, then $h(t) = (a+t)/(b+t)$ is increasing on $\mathbb{R}^+$. This completes the proof of (6.1).

Now let $u \in \mathscr{S}$, $u > 0$ everywhere, $\mu(\{u = \infty\}) = 0$ and $[\log(u/\pi u)]^- \in L^1(\mu)$. Let $u_n = \phi_n \circ u$; then $u_n \in \mathscr{U}$. Moreover, $u_n \to u$ and it easily follows that $\pi u_n \to \pi u$. By (6.1) and Fatou's lemma,

$$I(\mu) \ge \liminf_{n\to\infty} \int \log\left(\frac{u_n}{\pi u_n}\right) d\mu \ge \int \liminf_{n\to\infty} \log\left(\frac{u_n}{\pi u_n}\right) d\mu = \int \log\left(\frac{u}{\pi u}\right) d\mu.$$

It follows that $I(\mu) \ge I'(\mu)$. □

For the next lemma, we recall ([12], page 8) that a set $F \in \mathscr{S}$ is *π-closed* if $\pi(x, F^c) = 0$ for all $x \in F$.

LEMMA 6.2. *Let F be π-closed, and assume: There exists $n \ge 1$ such that for all $x \in F^c$, $\pi^n(x, F^c) = 0$. Suppose $\mu(F^c) > 0$. Then $I(\mu) = \infty$.*

PROOF. Let $\alpha > 0$, $\beta > 0$ and define $u = \alpha I_F + \beta I_{F^c}$. Then for $k \ge 1$,

$$(\pi^k u)(x) = \alpha\pi^k(x, F) + \beta\pi^k(x, F^c).$$

Let $m = \min\{k: \mu(\{x: \pi^{k+1}(x, F^c) > 0\}) = 0\}$. Observe that m exists, since $\{x: \pi^n(x, F^c) > 0\}$ is empty, $m \le n-1$, and $\mu(\{x: \pi^m(x, F^c) > 0\}) > 0$. We have, since $(\pi^k u)(x) = \alpha$ for $x \in F$, $k \ge 1$,

$$\int \log\left(\frac{\pi^m u}{\pi(\pi^m u)}\right) d\mu = \int_{F^c} \log\left[\frac{\alpha\pi^m(x,F) + \beta\pi^m(x,F^c)}{\alpha}\right] d\mu.$$

Now clearly $\mu(F^c \cap \{x: \pi^m(x, F^c) > 0\}) > 0$. Taking $\alpha = 1$ and β arbitrarily large, it follows that $I(\mu) = \infty$. □

In the next result, P_μ and $\{X_j, j \geq 0\}$ will be as in Theorem 5.4. $\{L_n\}$ is the sequence of occupation times of the Markov chain $\{X_j\}$: For $\omega \in S^{\mathbb{N}}$,

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

THEOREM 6.3. *Suppose that $(S, \mathscr{S})$ is a Polish space with its Borel σ-algebra and that π is an irreducible Markov kernel on $(S, \mathscr{S})$. Suppose also that the following condition is satisfied: For every π-closed set F, there exists $k \geq 1$ such that $\pi^k(x, F^c) = 0$ for all $x \in F^c$. Then for every probability measure μ on $(S, \mathscr{S})$ and every weakly open set $G \subset \mathscr{M}_1^+(S)$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log P_\mu(\{\omega: L_n(\omega, \cdot) \in G\}) \geq -I(G),$$

where $I(G) = \inf_{\nu \in G} I(\nu)$.

PROOF. We apply Theorem 5.6 to the setup described following Definition 5.5. Let $f: S \to \mathscr{M}_1^+(S)$ be defined by

$$f(x) = \delta_x.$$

Then f is a continuous map and for all $\omega \in S^{\mathbb{N}}$,

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

By Theorem 5.6, for every $\nu \in G$,

$$\liminf_{n\to\infty} \frac{1}{n} \log P_\mu(\{\omega: L_n(\omega, \cdot) \in G\}) \geq -\lambda(\nu).$$

The result will follow if we can prove: For every $\nu \in \mathscr{M}_1^+(S)$,

$$\lambda(\nu) \leq I(\nu). \tag{6.3}$$

Let $g \in C_b(S)$, $K_g(x, A) = \int_A e^{g(y)}\pi(x, dy)$ and $\phi(g) = -\log R(K_g)$. Let $\rho < R(K_g)$ and let s be a small function ([12], page 15). Then if

$$u = \sum_0^\infty \rho^n K_g^n s,$$

$u < \infty$ on a π-closed set F ([12], Definition 3.2). Moreover, $u > 0$ everywhere by irreducibility. Let $v = e^g u$; then

$$\pi v = K_g u = \rho^{-1}(u - s).$$

It follows that $v, \pi v > 0$ everywhere and $v, \pi v < \infty$ on F. Next, on the set F,

$$\log\left(\frac{v}{\pi v}\right) = \log\left(\frac{e^g u}{\rho^{-1}(u - s)}\right) \geq g + \log \rho. \tag{6.4}$$

Let ν be such that $I(\nu) < \infty$. Then by Lemma 6.2, $\nu(F^c) = 0$. It follows from (6.4) that $[\log(v/\pi v)]^- \in L^1(\nu)$ and

$$\int \log\left(\frac{v}{\pi v}\right) d\nu \geq \int g\, d\nu + \log \rho.$$

By Lemma 6.1,

$$I(\nu) = I'(\nu) \geq \int g\,d\nu + \log \rho,$$

and therefore

$$I(\nu) \geq \int g\,d\nu - \phi(g).$$

Since g was arbitrarily chosen in $C_b(S)$ and $C_b(S) = (\mathscr{M}(S), \sigma(\mathscr{M}(S), C_b(S)))^*$, (6.3) follows. □

REMARKS. 1. The condition in Theorem 5.3 holds if there exists $\phi \in \mathscr{I}(\pi)$ such that $\pi(x,\cdot) \ll \phi$ for all $x \in S$. For, if F is π-closed, then $\phi(F^c) = 0$ ([12], Proposition 2.5) and, therefore, $\pi(x, F^c) = 0$ for all $x \in S$.

2. Another approach to large deviations for occupation times of Markov chains, different from that of [5] and that of the present paper, is presented in Stroock [17]. However, in this work a very strong assumption is imposed on the kernel π; as a consequence, the lower bounds are uniform with respect to the starting point.

7. Some remarks on the spectral radius of K_ξ and upper bounds. Let π be a Markov kernel on $(S, \mathscr{S})$. Let E be a locally convex Hausdorff topological vector space, and let $f\colon S \to E$ be measurable. Let $B(S)$ be the space of real-valued bounded measurable functions on S, with the uniform norm. Assume that for all $\xi \in E^*$,

$$\sup_{x\in S} K_\xi(x, S) < \infty.$$

Then K_ξ defines a bounded operator on $B(S)$, which we will still denote K_ξ. The spectral radius of K_ξ will be denoted $r(K_\xi)$ and we define $\psi(\xi) = \log r(K_\xi)$, $\xi \in E^*$. It is easily seen that

$$(7.1) \qquad \left(r(K_\xi)\right)^{-1} = \sup\left\{r \geq 0\colon \sum_0^\infty r^n \|K_\xi^n\| < \infty\right\}.$$

Suppose also that π is irreducible. Then by the definition of the convergence parameter ([12], pages 27 and 28) and (7.1), it is clear that $(r(K_\xi))^{-1} \leq R(K_\xi)$ and, therefore, for all $\xi \in E^*$,

$$(7.2) \qquad \phi(\xi) \leq \psi(\xi).$$

In general, there may be strict inequality in (7.2). In fact, this will be typically the case if S is infinite. Roughly speaking, the convergence parameter of a kernel K (when it exists, that is, for irreducible kernels) measures the pointwise growth of the powers K^n, while the inverse of the spectral radius measures the uniform growth of $\{K^n\}$. An example of an irreducible countable Markov matrix π_0 for which $(r(\pi_0))^{-1} = 1 < R(\pi_0)$ may be found in [16], page 208. Then we have $\phi(0) < 0 = \psi(0)$.

In [4], upper bounds for large deviations of vector-valued functionals of Markov chains were obtained in terms of ψ^*, the convex conjugate of ψ (defined on E), without the assumption of irreducibility. For lower bounds and under the assumption of irreducibility, it appears by the results of [11] and of the present paper that the proper rate function is given by ϕ^*. One important reason for this is the continuity property of the convergence parameter given by Theorem 2.1, a property not possessed by the inverse of the spectral radius. It is natural to ask if

$$\psi^* = \phi^*. \tag{7.3}$$

When π is irreducible, this equality means that the rate function for upper bounds in [4] coincides with the rate function for lower bounds in the present paper. Let us observe that (7.3) does not contradict possible strict inequality in (7.2), since for a proper convex function h: $E^* \to \overline{R}$ the equality $h^{**} = h$ holds if and only if h is $\sigma(E^*, E)$-lower semicontinuous ([6], Propositions 4.1 and 3.1). It follows from (7.2) that for all $x \in E$,

$$\psi^*(x) \le \phi^*(x). \tag{7.4}$$

In general, (7.3) is false. In fact, let π_0 be as before (so $S = \mathbb{N}$) and let f: $\mathbb{N} \to \mathbb{R}^d$ be bounded. We know by Lemma 4.1 that ϕ is a continuous, proper convex function on $(\mathbb{R}^d)^*$ and it is not difficult to show that ψ has the same properties. If (7.3) were true, then the duality theorem for conjugate functions would imply

$$\phi = \phi^{**} = \psi^{**} = \psi;$$

in particular, $\phi(0) = \psi(0)$, which is false.

One important situation in which (7.3) holds is the case of occupation times. We use the setup described following Definition 5.5. In this case we have: If π is a Markov kernel on a Polish space S and π is Feller, irreducible and satisfies the assumption in Theorem 6.3, then for all $\nu \in \mathscr{M}(S)$,

$$\psi^*(\nu) = \phi^*(\nu) \tag{7.5}$$

and if $\nu \in \mathscr{M}_1^+(S)$, then the common value is $I(\nu)$. In fact, by [4], page 562, or [17], page 134, we have $I(\nu) \le \psi^*(\nu)$ for any $\nu \in \mathscr{M}_1^+(S)$. On the other hand, by (6.3), $\phi^*(\nu) = \lambda(\nu) \le I(\nu)$. Using (7.4), it follows that (7.5) holds for $\nu \in \mathscr{M}_1^+(S)$. If $\nu \notin \mathscr{M}_1^+(S)$, then by [17], pages 133 and 134, we have $\psi^*(\nu) = \infty$, so in this case (7.5) again follows from (7.4).

In [11] upper bounds in terms of ϕ^* are obtained in the case f: $S \to \mathbb{R}^d$ (more generally, for Markov additive processes) under restrictive conditions.

Taking into account these observations, we close by stating two problems:

1. Find conditions under which $\psi^* = \phi^*$.
2. Prove a general upper bound result for vector-valued f and irreducible π in terms of ϕ^*, relaxing the assumptions in [11].

A quantity introduced in [17], page 143, is relevant to these questions, or variants thereof.

APPENDIX A.

The ergodic theorem for vector-valued functionals of a Markov chain. Given a Markov kernel π and a probability measure μ on $(S, \mathscr{S})$, let P_μ, $\{X_n, n \geq 0\}$ be as in Section 5.

DEFINITION A.1. π is *Harris-recurrent* if there exists a σ-finite measure $\varphi \neq 0$ on $(S, \mathscr{S})$ such that $B \in \mathscr{S}$, $\varphi(B) > 0$ implies $P_x(\{X_n \in B$ infinitely often$\}) = 1$ for all $x \in S$.

It is well known (see Revuz [13], Chapter 3, Section 2, or Nummelin [12], Sections 3.6 and 5.2—the notions of Harris recurrence in the two books are superficially different but in fact equivalent) that if π is Harris-recurrent, then it has a σ-finite invariant measure, unique up to multiplication by a positive scalar. Of course, if π is Harris-recurrent and has a finite invariant measure, then it has a unique invariant probability measure.

The following result is a particular case of Theorem 3.6 of Revuz [13], page 123.

LEMMA A.2. *Suppose π is Harris-recurrent and has an invariant probability measure γ. Then for every $g \in L^1(S, \mathscr{S}, \gamma)$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} g(X_j) = \int g \, d\gamma, \qquad P_\mu\text{-a.s.},$$

for any probability measure μ.

The next result generalizes Lemma A.2 to separable Banach spaces. The method of proof is well known (see [7], page 131 and [8]).

THEOREM A.3. *Let π be as in Lemma* A.2 *and let E be a separable Banach space. Then the conclusion of Lemma* A.2 *holds for every $g \in L^1(S, \mathscr{S}, \gamma; E)$.*

PROOF. First assume that g is an E-valued simple function, $g = \sum_{i=1}^k x_i I_{A_i}$, $A_i \in \mathscr{S}$, $x_i \in E$, $i = 1, \ldots, k$. Then by Lemma A.2,

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{n-1} g(X_j) &= \sum_{i=1}^k x_i \left[\frac{1}{n} \sum_{j=0}^{n-1} I_{A_i}(X_j) \right] \\ &\to \sum_{i=1}^k x_i \int I_{A_i} \, d\gamma = \int g \, d\gamma, \qquad P_\mu\text{-a.s.} \end{aligned} \tag{A.1}$$

Now let $g \in L^1(S, \mathscr{S}, \gamma; E)$. By [10], pages 101–102, there exists a sequence $\{g_k\}$ of E-valued simple functions such that for $k \geq 1$,

$$\|g_k\| \leq \|g\|, \qquad g_k \to g \quad \text{pointwise and} \quad \int \|g - g_k\| \, d\gamma < k^{-1}.$$

By Lemma A.2, for each $k \geq 1$,

$$\frac{1}{n}\sum_{j=0}^{n-1}\|(g-g_k)(X_j)\| \to \int\|g-g_k\|\,d\gamma, \qquad P_\mu\text{-a.s.} \tag{A.2}$$

By the triangle inequality,

$$\left\|\frac{1}{n}\sum_{j=0}^{n-1} g(X_j) - \int g\,d\gamma\right\| \leq \frac{1}{n}\sum_{j=0}^{n-1}\left\|(g-g_k)(X_j)\right\| + \left\|\frac{1}{n}\sum_{j=0}^{n-1} g_k(X_j) - \int g_k\,d\gamma\right\| + \left\|\int g_k\,d\gamma - \int g\,d\gamma\right\|.$$

Fixing k and letting $n \to \infty$, by (A.1) and (A.2) we have

$$\limsup_{n\to\infty}\left\|\frac{1}{n}\sum_{j=0}^{n-1} g(X_j) - \int g\,d\gamma\right\| \leq 2k^{-1}, \qquad P_\mu\text{-a.s.}$$

Since k is arbitrary, the result follows. □

APPENDIX B.

A property of the Fenchel transform. Given a dual system (V, W) of real vector spaces, the *Fenchel transform* is the map which assigns to each $g\colon V \to \overline{\mathbb{R}}$ its convex conjugate $g^*\colon W \to \overline{\mathbb{R}}$, defined by

$$g^*(w) = \sup_{v\in V}\left[\langle v, w\rangle - g(v)\right], \qquad w \in W. \tag{B.1}$$

Mosco [9] (see also Attouch [1], Chapter 3) has established a continuity property of the Fenchel transform in the case when V is a reflexive Banach space and W is its dual space; this result, however, is not suitable for our needs in Lemma 5.1. We shall prove in the following text a continuity property which is well adapted to our purposes; in this result, given a separable Banach space E, we take $V = E^*$ endowed with the w^*-topology and $W = E$. Our proof consists in suitable variations of arguments in [1], Section 3.2.

LEMMA B.1. *Let E be a Banach space. For given $a \in \mathbb{R}$, $b > 0$, $c > 0$, define $g\colon E^* \to \mathbb{R}$ by*

$$g(\xi) = a - b\|\xi\| + (c/2)\|\xi\|^2, \qquad \xi \in E^*.$$

Then

$$g^*(x) = -a + \frac{(\|x\| + b)^2}{2c}, \qquad x \in E.$$

PROOF. Fix $x \in E$. Then for $\xi \in E^*$,

$$\begin{aligned}\langle \xi, x\rangle - g(\xi) &\leq \|\xi\|\,\|x\| - a + b\|\xi\| - (c/2)\|\xi\|^2\\ &= -a + (\|x\| + b)\|\xi\| - (c/2)\|\xi\|^2.\end{aligned}$$

Hence

$$\text{(B.2)}\qquad \begin{aligned} g^*(x) &= \sup_{\xi\in E^*}[\langle \xi, x\rangle - g(\xi)] \le \sup_{t\in\mathbb{R}^+}\left[-a + (\|x\| + b)t - (c/2)t^2\right] \\ &= -a + \frac{(\|x\| + b)^2}{2c}. \end{aligned}$$

On the other hand, by the Hahn–Banach theorem there exists $\xi \in E^*$ such that $\|\xi\| = 1$ and $\langle \xi, x\rangle = \|x\|$. Let $t = (\|x\| + b)/c$. Then a simple calculation shows that

$$\text{(B.3)}\qquad \langle t\xi, x\rangle - g(t\xi) = -a + \frac{(\|x\| + b)^2}{2c}.$$

The assertion follows from (B.2) and (B.3). □

LEMMA B.2. *Let E be a Banach space. Let g be a convex, proper, w^*-lower semicontinuous function defined on E^*. Define, for $c > 0$,*

$$g_c(\xi) = g(\xi) + \frac{c}{2}\|\xi\|^2.$$

Then

$$g_c^*(x) = \inf_{y\in E}\left\{g^*(y) + \frac{1}{2c}\|x - y\|^2\right\}, \qquad x \in E.$$

PROOF. Let

$$h(x) = \inf_{y\in E}\left\{g^*(y) + \frac{1}{2c}\|x - y\|^2\right\}, \qquad x \in E.$$

The function g^* is convex, proper and lower semicontinuous (see [6], Chapter 1, Sections 3 and 4). By [1], Proposition 3.3, page 266, for all $\xi \in E^*$,

$$h^*(\xi) = g^{**}(\xi) + (c/2)\|\xi\|^2.$$

The assumptions on g and the duality theorem for the Fenchel transform (see [6], Proposition 4.1, page 18 and Proposition 3.1, page 14) imply that $g^{**} = g$. Hence $h^* = g_c$ and therefore $g_c^* = h^{**}$. Now the conclusion will follow by another application of the duality theorem if we can prove that h is convex, proper, lower semicontinuous.

(i) h is convex. In fact, let $x, y \in E$, $\alpha > 0$, $\beta > 0$ with $\alpha + \beta = 1$. Then for all $v, w \in E$,

$$\begin{aligned} &g^*(\alpha v + \beta w) + \frac{1}{2c}\|\alpha x + \beta y - (\alpha v + \beta w)\|^2 \\ &\quad\le \alpha g^*(v) + \beta g^*(w) + \frac{1}{2c}\left[\alpha\|x - v\|^2 + \beta\|y - w\|^2\right] \\ &\quad\le \alpha\left[g^*(v) + \frac{1}{2c}\|x - v\|^2\right] + \beta\left[g^*(w) + \frac{1}{2c}\|y - w\|^2\right] \end{aligned}$$

and, therefore, $h(\alpha x + \beta y) \le \alpha h(x) + \beta h(y)$.

(ii) h is proper. In fact, if $g(\xi_0) < \infty$, then for any $x, y \in E$,

$$g^*(y) + \frac{1}{2c}\|x - y\|^2$$

$$\geq \langle \xi_0, y \rangle - g(\xi_0) + \frac{1}{2c}\|x - y\|^2 \tag{B.4}$$

$$\geq -g(\xi_0) + \langle \xi_0, x \rangle - \|\xi_0\| \, \|x - y\| + \frac{1}{2c}\|x - y\|^2,$$

which implies $h(x) > -\infty$. Also $g^*(y) < \infty$ for some $y \in E$ and, therefore, $h(x) < \infty$ for all $x \in E$.

(iii) h is lower semicontinuous. In fact, assume $x_n \to x$ in E. If $\liminf_{n \to \infty} h(x_n) = \infty$, there is nothing to prove. Otherwise, we may assume that $L = \sup_n h(x_n) < \infty$. Choose $\{y_n\} \subset E$ so that

$$g^*(y_n) + \frac{1}{2c}\|x_n - y_n\|^2 \leq h(x_n) + \varepsilon_n, \tag{B.5}$$

where $\varepsilon_n \downarrow 0$.

Now (B.4) and (B.5) imply

$$a = \sup_n \|x_n - y_n\| < \infty.$$

Next, for any $\varepsilon > 0$,

$$h(x) \leq g^*(y_n) + \frac{1}{2c}\|x - y_n\|^2 \leq g^*(y_n) + \frac{1}{2c}(\|x - x_n\| + \|x_n - y_n\|)^2$$

$$\leq g^*(y_n) + \frac{1}{2c}\|x_n - y_n\|^2 + \frac{\varepsilon^2}{2c}\|x_n - y_n\|^2 + \frac{1}{2c}\left(1 + \frac{1}{\varepsilon^2}\right)\|x - x_n\|^2$$

$$\leq h(x_n) + \varepsilon_n + \frac{\varepsilon^2}{2c}a + \frac{1}{2c}\left(1 + \frac{1}{\varepsilon^2}\right)\|x - x_n\|^2.$$

Therefore $h(x) \leq \liminf_{n \to \infty} h(x_n) + (\varepsilon^2/2c)a$. Since ε is arbitrary, the conclusion follows. □

THEOREM B.3. *Let E be a separable Banach space. Let g_n, $n \geq 1$, and g be functions from E^* into $\overline{\mathbb{R}}$, such that:*

(i) *For $n \geq 1$, g_n is convex, proper, w^*-lower semicontinuous (w^*-l.s.c.).*
(ii) *$g_n \uparrow g$ pointwise and g is proper.*
(iii) *There exist constants $a \in \mathbb{R}$, $b > 0$ such that for all $\xi \in E^*$,*

$$g_1(\xi) \geq a - b\|\xi\|.$$

Then for every $x \in E$, there exists a sequence $\{x_n\} \subset E$ such that $x_n \to x$ and $\limsup_{n \to \infty} g_n^*(x_n) \leq g^*(x)$.

PROOF.

CLAIM 1. Assumptions (i) and (ii) imply: If $\xi = w^*\text{-}\lim_n \xi_n$, then

$$g(\xi) \leq \liminf_{n \to \infty} g_n(\xi_n).$$

To prove this claim, let $\mathscr{N}$ be the w^*-neighborhood system of ξ. Then

$$\begin{aligned} g(\xi) &= \sup_n g_n(\xi) \\ &= \sup_n \left[\sup_{V\in\mathscr{N}} \inf_{\eta\in V} g_n(\eta) \right] \quad \text{(by the } w^*\text{-l.s.c. of } g_n) \\ &= \sup_{V\in\mathscr{N}} \sup_n \inf_{\eta\in V} g_n(\eta) \\ &= \sup_{V\in\mathscr{N}} \liminf_{n\to\infty} \inf_{\eta\in V} g_n(\eta). \end{aligned}$$

If $g(\xi) < \infty$, then given $\varepsilon > 0$, there exists $V_\varepsilon \in \mathscr{N}$ such that

$$g(\xi) \le \liminf_{n\to\infty} \inf_{\eta\in V_\varepsilon} g_n(\eta) + \varepsilon. \tag{B.6}$$

Choose n_0 such that $\xi_n \in V_\varepsilon$ for $n \ge n_0$. Then for $n \ge n_0$,

$$\inf_{\eta\in V_\varepsilon} g_n(\eta) \le g_n(\xi_n)$$

and, therefore,

$$\liminf_{n\to\infty} \inf_{\eta\in V_\varepsilon} g_n(\eta) \le \liminf_{n\to\infty} g_n(\xi_n),$$

which together with (B.6) yields

$$g(\xi) \le \liminf_{n\to\infty} g_n(\xi_n) + \varepsilon.$$

But ε is arbitrary. This proves the claim in the case $g(\xi) < \infty$. The case $g(\xi) = \infty$ is similar.

CLAIM 2. For fixed $c > 0$, define on E^*

$$g_{c,n}(\xi) = g_n(\xi) + \frac{c}{2}\|\xi\|^2, \qquad g_c(\xi) = g(\xi) + \frac{c}{2}\|\xi\|^2.$$

Then for all $x \in E$,

$$\lim_n g^*_{c,n}(x) = g^*_c(x).$$

To prove this claim, we first observe that, since $g_n \uparrow g$, we have $g_{c,n} \uparrow g_c$ and $\{g^*_{c,n}\}$ is a decreasing sequence of functions on E. From assumption (iii) and Lemma B.1 it follows that for all $x \in E$,

$$g^*_{c,n}(x) \le g^*_{c,1}(x) \le -a + \frac{(\|x\| + b)^2}{2c} < \infty.$$

Also, from the fact that g is proper it follows that for all $x \in E$,

$$g^*_{c,n}(x) \ge g^*_c(x) > -\infty.$$

Fix $x \in E$ and let $l = \lim_n g^*_{c,n}(x)$. Then $l \in \mathbb{R}$ and $l \ge g^*_c(x)$. To prove the opposite inequality, let $\{\xi_n\} \subset E^*$ be such that

$$\langle \xi_n, x\rangle - g_{c,n}(\xi_n) \ge g^*_{c,n}(x) - \varepsilon_n,$$

where $\varepsilon_n \downarrow 0$; then $\langle \xi_n, x\rangle - g_{c,n}(\xi_n) \to l$. By assumption (iii), for $n \geq 1$,

$$\langle \xi_n, x\rangle - g_{c,n}(\xi_n) \leq \langle \xi_n, x\rangle - \left[a - b\|\xi_n\| + \frac{c}{2}\|\xi_n\|^2\right]$$

and, therefore, $\{\xi_n\}$ is bounded. By the Banach–Alaoglu theorem (and the separability of E), there exist a subsequence $\{\xi_{n_k}\}$ of $\{\xi_n\}$ and a point $\xi \in E^*$ such that $w^*\text{-}\lim_k \xi_{n_k} = \xi$. Then by Claim 1,

$$\begin{aligned} g_c(\xi) &\leq \liminf_{k\to\infty} g_{n_k}(\xi_{n_k}) + \frac{c}{2}\liminf_{k\to\infty}\|\xi_{n_k}\|^2 \\ &\leq \liminf_{k\to\infty} g_{c,n_k}(\xi_{n_k}) \end{aligned}$$

and, therefore,

$$\begin{aligned} l &= \lim_k \left[\langle \xi_{n_k}, x\rangle - g_{c,n_k}(\xi_{n_k})\right] \\ &\leq \liminf_{k\to\infty}\langle \xi_{n_k}, x\rangle + \limsup_{k\to\infty}\left[-g_{c,n_k}(\xi_{n_k})\right] \\ &\leq \langle \xi, x\rangle - g_c(\xi) \\ &\leq g_c^*(x), \quad \text{proving Claim 2.} \end{aligned}$$

Next

$$\begin{aligned} \sup_{c>0} g_c^*(x) &= \sup_{c>0}\sup_{\xi\in E^*}\left[\langle \xi, x\rangle - g_c(\xi)\right] \\ &= \sup_{\xi\in E^*}\sup_{c>0}\left[\langle \xi, x\rangle - g(\xi) - \frac{c}{2}\|\xi\|^2\right] \qquad \text{(B.7)} \\ &= g^*(x). \end{aligned}$$

By Claim 2 and (B.7), we have

$$\limsup_{c\to 0}\lim_n g_{c,n}^*(x) \leq g^*(x).$$

By a diagonalization lemma ([1], Corollary 1.16, page 33), there exists a decreasing map $\lambda\colon \mathbb{N} \to \mathbb{R}^+$, such that $\lim_n \lambda(n) = 0$ and

$$\limsup_{n\to\infty} g_{\lambda(n),n}^*(x) \leq g^*(x). \qquad \text{(B.8)}$$

Let $\{x_n\} \subset B$ be such that

$$g_n^*(x_n) + \frac{1}{2\lambda(n)}\|x - x_n\|^2 \leq \inf_{y\in B}\left\{g_n^*(y) + \frac{1}{2\lambda(n)}\|x - y\|^2\right\} + \varepsilon_n, \qquad \text{(B.9)}$$

where $\varepsilon_n \downarrow 0$. Then by Lemma B.2, (B.8) and (B.9),

$$\limsup_{n\to\infty} g_n^*(x_n) \leq \limsup_{n\to\infty}\left[g_{\lambda(n),n}^*(x) + \varepsilon_n\right] \leq g^*(x). \qquad \text{(B.10)}$$

Assume now that $g^*(x) < \infty$. We will show that $\|x - x_n\| \to 0$. In fact, let ξ_0 be such that $g(\xi_0) < \infty$. Then

$$g_n^*(x_n) \geq g^*(x_n) \geq \langle \xi_0, x_n\rangle - g(\xi_0)$$

and by (B.9) and (B.10),

$$\limsup_{n\to\infty}\left[\langle \xi_0, x_n\rangle - g(\xi_0) + \frac{1}{2\lambda(n)}\|x - x_n\|^2\right] \le g^*(x),$$

which implies $\|x - x_n\| \to 0$. On the other hand, if $g^*(x) = \infty$, there is nothing to prove. □

REFERENCES

[1] ATTOUCH, H. (1984). *Variational Convergence for Functions and Operators*. Pitman, Boston.

[2] AUBIN, J.-P. and EKELAND, I. (1984). *Applied Nonlinear Analysis*. Wiley, New York.

[3] BRONDSTED, A. and ROCKAFELLAR, R. T. (1965). On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.* **16** 605–611.

[4] DE ACOSTA, A. (1985). Upper bounds for large deviations of dependent random vectors. *Z. Wahrsch. verw. Gebiete* **69** 551–565.

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

[6] EKELAND, I. and TEMAM, R. (1976). *Convex Analysis and Variational Problems*. North-Holland, Amsterdam.

[7] HOFFMANN-JORGENSEN, J. (1976). *Probability in Banach Space. Lecture Notes in Math.* **598**. Springer, Berlin.

[8] LANDERS, D. and ROGGE, L. (1978). An ergodic theorem for Fréchet-valued random variables. *Proc. Amer. Math. Soc.* **72** 49–53.

[9] MOSCO, U. (1971). On the continuity of the Young–Fenchel transform. *J. Math. Anal. Appl.* **35** 518–535.

[10] NEVEU, J. (1972). *Martingales á Temps Discret*. Masson, Paris.

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

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

[13] REVUZ, D. (1975). *Markov Chains*. North-Holland, Amsterdam.

[14] ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press, Princeton, N.J.

[15] ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. MacMillan, New York.

[16] SENETA, E. (1981). *Non-negative Matrices and Markov Chains*, 2nd ed. Springer, Berlin.

[17] STROOCK, D. W. (1984). *An Introduction to the Theory of Large Deviations*. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
CASE WESTERN RESERVE UNIVERSITY
CLEVELAND, OHIO 44106