Z. Wahrscheinlichkeitstheorie verw. Gebiete
69, 551–565 (1985)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Upper Bounds for Large Deviations of Dependent Random Vectors*

A. de Acosta

Department of Mathematics and Statistics, Case Western Reserve University, Cleveland, Ohio 44106, USA

1. Introduction

In this paper we obtain upper bounds for large deviations for certain classes of dependent random vectors. One of the situations we study is as follows.

Let $\{X_j, j \geqq 0\}$ be a Markov chain with state space S and transition probability π. Let E be a topological vector space and $f\colon S \to E$. We obtain upper bounds of the type

$$\limsup_n n^{-1} \log P\left\{n^{-1} \sum_{j=0}^{n-1} f(X_j) \in F\right\} \leqq -\Lambda(F) \tag{1.1}$$

where F is a closed subset of E,

Λ is a set functional associated by convex duality to ϕ, where for $\xi \in E'$, the dual of E, $\phi(\xi) = \log r(T_\xi)$, and $r(T_\xi)$ is the spectral radius of a certain operator naturally associated to ξ, π and f.

This extends some of the work on large deviations by Donsker and Varadhan [6].

Another situation that we study is that of sums of exchangeable random vectors.

In Sect. 2 we present a general result on upper bounds for large deviations of dependent random vectors, slightly extending some work of Ellis [8].

Section 3 contains an integrability theorem which provides the basis for the extension of upper bounds from compact sets to closed sets in Sects. 4 and 5. Even in the case of independent identically distributed random vectors Theorem 3.1 (with the family $\{\mu_\alpha\}$ reduced to one measure) simplifies the methods in the literature ([6, 1]) and appears to have independent technical interest.

In Sect. 4 we prove the upper bound (1.1), slightly strengthened, as an application of Theorem 2.1. The content of Theorem 4.2 is that under very weak restrictions on π and f a natural upper bound of the form (1.1) exists for

* This paper is a revised version of "Upper Bounds for Large Deviations of Vector-Valued Functionals of a Markov Chain", Department of Mathematics and Statistics, Case Western Reserve University preprint, January 1984

compact sets and, if suitable integrability and tightness conditions are imposed, for closed sets as well. Then we show that some important results on upper bounds for large deviations of infinite-dimensional random vectors recently proved in the literature may be obtained as corollaries of Theorem 4.2. We consider the case of independent, identically distributed random vectors taking values in a separable Banach space (Donsker and Varadhan [6]; Bahadur and Zabell [2]; also Azencott [1]) and the case of occupation times of a Markov chain (Donsker and Varadhan [6]). Our approach has some features in common with Ch. 7 of the interesting very recent book [15], although the points of view are different (this book appeared half a year after the present work had been submitted).

In Sect. 5 we prove a result on upper bounds for sums of exchangeable random vectors.

To close this introduction, we remark that we have recently obtained results on lower bounds in the framework of Sects. 2 and 4. At present, however, our results for lower bounds are less simple and general than the upper bound result.

2. Upper Bounds for Large Deviations of Dependent Random Vectors

In this section we prove a general result on upper bounds for large deviations of dependent random vectors under an assumption on the limiting behavior of their Laplace transform. The case of compact sets has been discussed in Ellis [8], on the basis of an idea of Gärtner [9]; we give a somewhat strengthened version using their approach (for other related references see [8]). The condition (2.3) below, which makes it possible to pass from compact sets to closed sets, is used in [6] and has been isolated by Azencott [1] in the case of partial sums of independent identically distributed random vectors. A technical observation perhaps worth emphasizing is that upper bounds for large deviations of random vectors $\{Y_n\}$ depend only on a limiting *inequality* for the normalized logarithms of the Laplace transforms of $\{Y_n\}$.

Let E be a Hausdorff topological vector space, endowed with its Borel σ-algebra. Let E' be the dual space of E; the weak topology induced on E by E' will be denoted $\sigma(E,E')$. Given a function ϕ: $E'\to\bar{\mathbb{R}}$, its convex conjugate λ_ϕ is defined by

$$\lambda_\phi(x)=\sup_{\xi\in E'}[\langle\xi,x\rangle-\phi(\xi)] \qquad (x\in E)$$

and the *Cramér functional* Λ_ϕ of ϕ is defined by

$$\Lambda_\phi(A)=\inf_{x\in A}\lambda_\phi(x) \qquad (A\subset E).$$

Theorem 2.1. *Let $\{P_\alpha,\alpha\in I\}$ be a family of probability measures on a measurable space $(\Omega,\mathscr{A})$ and for each $\alpha\in I$, let E_α be the P_α-expectation functional. Let $\{Y_n\}$ be a sequence of E-valued random vectors, defined on Ω. Assume: for every $\xi\in E'$,*

$$\limsup_n n^{-1} \log \sup_{\alpha\in I} E_\alpha \exp\langle \xi, Y_n\rangle \leqq \phi(\xi) \tag{2.1}$$

for a certain ϕ: $E' \to \mathbb{R}$. *Then*

(i) *for every* $\sigma(E, E')$-*compact set* $F \subset E$,

$$\limsup_n n^{-1} \log \sup_{\alpha\in I} P_\alpha\{n^{-1} Y_n \in F\} \leqq -\Lambda_\phi(F); \tag{2.2}$$

(ii) *if, furthermore, the following condition holds: for every* $a>0$, *there exist a compact set* $K_a \subset E$ *and* $n_0 \in \mathbb{N}$ *such that*

$$\sup_{\alpha\in I} P_\alpha\{n^{-1} Y_n \in K_a^c\} \leqq e^{-na} \qquad (n \geqq n_0), \tag{2.3}$$

then (2.2) *holds also for every closed set* F.

Proof. Let $\lambda = \lambda_\phi$, $\Lambda = \Lambda_\phi$. We have several cases:

(1) $\Lambda(F) \leqq 0$. This case is trivial.

(2) $0 < \Lambda(F) < \infty$. Let $\varepsilon > 0$ and introduce

$$H(\xi) = \{x \in E: \langle \xi, x\rangle - \phi(\xi) > \Lambda(F) - \varepsilon\} \qquad (\xi \in E').$$

Then

$$F \subset \{x \in E: \lambda(x) > \Lambda(F) - \varepsilon\} = \bigcup_{\xi\in E'} H(\xi).$$

Since F is $\sigma(E, E')$-compact, there exist $\xi_1, \ldots, \xi_k$ in E' such that

$$F \subset \bigcup_{i=1}^{k} H(\xi_i).$$

Now $P_\alpha\{Y_n \in nF\} \leqq \sum_{i=1}^{k} P_\alpha\{Y_n \in nH(\xi_i)\}$

$$= \sum_{i=1}^{k} P_\alpha\{\langle \xi_i, Y_n\rangle > n(\phi(\xi_i) + b)\},$$

where $b = \Lambda(F) - \varepsilon$, $\leqq \sum_{i=1}^{k} e^{-n(\phi(\xi_i)+b)} E_\alpha \exp\langle \xi_i, Y_n\rangle$

$$\leqq k \exp\{-n(\inf_i \{\phi(\xi_i) - n^{-1} \log E_\alpha \exp\langle \xi_i, Y_n\rangle\} + b)\}.$$

By assumption (2.1),

$$\limsup_n n^{-1} \log \sup_{\alpha\in I} P_\alpha\{Y_n \in nF\} \leqq -b = -\Lambda(F) + \varepsilon.$$

Since ε is arbitrary, (2.2) follows.

(3) $\Lambda(F) = \infty$. In this case, fix $m > 0$ and consider

$$L(\xi) = \{x \in E: \langle \xi, x\rangle - \phi(\xi) > m\}.$$

Then $F \subset \bigcup_{\xi\in E'} L(\xi)$, and proceeding as in (2) one obtains

$$\limsup_n n^{-1} \log \sup_{\alpha\in I} P_\alpha\{Y_n \in nF\} \leqq -m.$$

Since m is arbitrary, (2.2) again follows. This completes the proof of (i).

To prove (ii) let $a>0$ and let K_a be a compact set satisfying (2.3). Then for every closed set F

$$\begin{aligned} P_\alpha\{n^{-1}Y_n\in F\} &= P_\alpha\{n^{-1}Y_n\in K_a\cap F\}+P_\alpha\{n^{-1}Y_n\in K_a^c\cap F\} \\ &\leqq 2\max(P_\alpha\{n^{-1}Y_n\in K_a\cap F\}, P_\alpha\{n^{-1}Y_n\in K_a^c\}) \end{aligned}$$

and since $K_a\cap F$ is $\sigma(E,E')$-compact we have by (i) and (2.3)

$$\begin{aligned} \limsup_n n^{-1}\log\sup_{\alpha\in I} P_\alpha\{n^{-1}Y_n\in F\} &\leqq \max\{-\Lambda(K_a\cap F), -a\} \\ &\leqq \max\{-\Lambda(F), -a\}. \end{aligned}$$

Since a is arbitrary, (2.2) follows. □

Remark. As observed by Ellis [8] the argument of Theorem 2.1 is still valid if $\{n\}$ is replaced by a positive sequence $\{a_n\}$ with $\lim_n a_n=\infty$.

3. An Integrability Theorem

A subset A of a vector space E is *positively balanced* if $\lambda x\in A$ whenever $x\in A$ and $\lambda\in[0,1]$. Given a convex, positively balanced set $A\subset E$, its *Minkowski functional* q_A is defined by

$$q_A(x)=\inf\{\lambda>0: x\in\lambda A\} \qquad (x\in E)$$

(with the customary convention: $\inf\phi=+\infty$). Then $q_A(x)<\infty$ if and only if $x\in\bigcup_n(nA)$, and q_A is subadditive and positively homogeneous.

Theorem 3.1. *Let E be a Hausdorff locally convex topological vector space, and let $E_0\subset E$ be a convex, positively balanced subset of E such that E_0 with the relative topology is Polish. Let p: $E_0\to\mathbb{R}^+$ be a measurable function such that*

(a) *p is subadditive and positively homogeneous,*

(b) *for every neighborhood V of 0 in E there exists $\varepsilon>0$ such that $\{x\in E_0: p(x)<\varepsilon\}\subset V\cap E_0$.*

Let $\{\mu_\alpha, \alpha\in I\}$ be a family of probability measures on E such that

(1) *$\mu_\alpha(E_0)=1$ for all $\alpha\in I$ and $\{\mu_\alpha,\alpha\in I\}$ is tight,*

(2) *for every $t>0$,*

$$\sup_{\alpha\in I}\int\exp(tp)\,d\mu_\alpha<\infty.$$

Then there exists a compact, convex, positively balanced subset K of E_0 such that

$$\sup_{\alpha\in I}\int\exp(q_K)\,d\mu_\alpha<\infty.$$

For the proof of Theorem 3.1 we need

Lemma 3.2. *Let $\tau_\alpha(t)=\mu_\alpha(\{x: p(x)>t\})(\alpha\in I, t>0)$. Then*

$$\lim_{t\to\infty}\sup_{\alpha\in I}(\tau_\alpha(t))^{1/t}=0.$$

Proof. Given $\varepsilon>0$, choose $a>0$ so that $e^{-a}<\varepsilon$. Then for all $t>0$,

$$\begin{aligned}\tau_\alpha(t) &\leqq e^{-at}\int \exp(ap)\,d\mu_\alpha\\ &\leqq e^{-at} M_a \quad \text{(independent of } \alpha \text{ by (2)).}\end{aligned}$$

Therefore

$$\sup_{\alpha\in I}(\tau_\alpha(t))^{1/t} \leqq e^{-a} M_a^{1/t},$$

$$\limsup_{t\to\infty}\sup_{\alpha\in I}(\tau_\alpha(t))^{1/t} \leqq e^{-a}<\varepsilon. \quad \square$$

Proof of Theorem 3.1. We will prove the following statement, which clearly implies the conclusion: for every $\beta\in(0,1)$, there exists $c>0$ and a compact, convex, positively balanced set $K\subset E_0$ such that

$$\mu_\alpha(\{x: q_K(x)>t\})\leqq c\beta^t \quad \text{for all } \alpha, \text{ all } t\geqq 1. \tag{3.1}$$

Choose $\beta\in(0,1)$. Set

$$t_{m,\alpha}=\inf\{t>0: \tau_\alpha(t)<\beta^m\};$$

it follows that $\tau_\alpha(t_{m,\alpha})\leqq\beta^m$. Let

$$B_{m,\alpha}=\{x\in E_0: p(x)\leqq t_{m,\alpha}\}.$$

By (1), there exists a compact set $K_m\subset E_0$ such that

$$\mu_\alpha(K_m^c)<\beta^m \quad \text{for all } \alpha;$$

by the assumption on E_0, we may assume that K_m is convex and positively balanced (see e.g. [14], p. 50). We may also assume $K_m\subset K_{m+1}$ for all m. Now let

$$A_m=\bigcup_{\alpha\in I} m^{-1}(K_m\cap B_{m,\alpha}),$$

$$K=\text{closed convex positively balanced hull of } (\bigcup_m A_m).$$

We claim that K is compact. To prove this, we first observe that

$$d_m=m^{-1}(\sup_{\alpha\in I} t_{m,\alpha})\to 0 \quad \text{as} \quad m\to\infty. \tag{3.2}$$

In fact, given $\varepsilon>0$, choose m_0 so that $m_0^{-1/2}<\varepsilon/2$ and for all α, all $t>m_0^{1/2}$,

$$\log\beta/(t^{-1}\log\tau_\alpha(t))<\varepsilon/2;$$

this is possible by Lemma 3.2. Let $m>m_0$. If $t_{m,\alpha}/2>m^{1/2}$, we have

$$\tau_\alpha(t_{m,\alpha}/2)\geqq\beta^m,$$

$$(t_{m,\alpha}/2)\log\tau_\alpha(t_{m,\alpha}/2)\geqq(t_{m,\alpha}/2)m\log\beta,$$

$$\frac{t_{m,\alpha}}{2m} \leqq \frac{\log \beta}{\left(\frac{2}{t_{m,\alpha}}\right) \log \tau_\alpha(t_{m,\alpha}/2)} < \varepsilon/2,$$

$$\frac{t_{m,\alpha}}{m} < \varepsilon.$$

On the other hand, if $t_{m,\alpha}/2 \leqq m^{1/2}$, then

$$\frac{t_{m,\alpha}}{m} \leqq \frac{2}{m^{1/2}} < \frac{2}{m_0^{1/2}} < \varepsilon.$$

Thus for $m > m_0$, we have $(\sup_{\alpha \in I} t_{m,\alpha})/m < \varepsilon$, proving (3.2). It follows that $(\bigcup_m A_m)$ is totally bounded; for, let V, ε be as in assumption (b). Choose m_0 so that $d_m < \varepsilon$ for $m \geqq m_0$. Then

$$\begin{aligned}
\bigcup_m A_m &= (\bigcup_{m \leqq m_0} A_m) \cup (\bigcup_{m > m_0} A_m) \\
&\subset K_{m_0} \cup \{x \in E_0 : p(x) \leqq \varepsilon\} \\
&\subset K_{m_0} \cup (V \cap E_0),
\end{aligned}$$

proving the total boundedness of $(\bigcup_m A_m)$. By the assumption on E_0, it follows that K is compact (see e.g. [14], p. 50), establishing the claim.

To conclude the proof we verify (3.1). Given $t \geqq 1$, let $m = [t]$. Then for all $\alpha \in I$,

$$\begin{aligned}
\mu_\alpha(\{x : q_K(x) > t\}) &= \mu_\alpha((tK)^c) \\
&\leqq \mu_\alpha((mK)^c) \\
&\leqq \mu_\alpha((K_m \cap B_{m,\alpha})^c) \\
&\leqq \mu_\alpha(K_m^c) + \tau_\alpha(t_{m,\alpha}) \\
&\leqq \beta^m + \beta^m = 2\beta^m,
\end{aligned}$$

and since $\beta^{m+1} < \beta^t$ we have

$$\mu_\alpha(\{x : q_K(x) > t\}) \leqq (2\beta^{-1})\beta^t. \quad \square$$

The following example, due to J. Rosinski [13], shows that the assumptions of Theorem 3.1 cannot be easily relaxed.

Example 3.2. For every infinite dimensional normed linear space E, there exists an E-valued random vector X such that $E(\exp \|X\|) < \infty$ but $E(\exp q_K(X)) = \infty$ for every compact, convex, balanced set K.

Proof. By the Riesz lemma (see e.g. [7], p. 578) there exists a sequence $\{x_n\} \subset E$ such that $\|x_n\| = 1$ for all n and $\|x_n - x_k\| > 1/2$ for $n \neq k$. Let $a_n = \log n$, $p_n = c n^{-3}$, where c is chosen so that $\sum_{n=1}^{\infty} p_n = 1$.

Let X be an E-valued random vector such that:

$$P\{X = a_n x_n\} = p_n.$$

An easy computation shows: $E(\exp\|X\|)<\infty$. On the other hand, the compactness of K implies $\lim_n q_K(x_n)=\infty$ and it easily follows that $E(\exp q_K(X)) =\infty$. $\square$

4. Vector Valued Functionals of a Markov Chain

We shall work with the canonical version of a Markov chain. That is, we take the underlying measurable space to be $S^{\mathbb{N}}$ with the product σ-algebra, where $(S,\mathscr{S})$ is a given measurable space, and $\{X_j, j\geqq 0\}$ are the coordinate maps; we denote by $\mathfrak{F}_k$ the σ-algebra generated by $X_0,\ldots,X_k$. θ is the shift operator on $S^{\mathbb{N}}$; that is, $\theta((x_j)_{j\geqq 0})=(x_j)_{j\geqq 1}$. The Markovian probability measure on $S^{\mathbb{N}}$ determined by the transition probability π and the initial distribution μ will be denoted P_μ; we write $P_x=P_{\delta_x}$ for $x\in S$ and E_μ for the P_μ-expectation functional.

We denote $B(S)$ the Banach space of real-valued bounded measurable functions defined on S, with the supremum norm $\|\cdot\|_\infty$.

Theorem 4.1. *Let M be a family of probability measures on S. Let E be a Hausdorff locally convex topological vector space and let $f\colon S\to E$ be a measurable map such that for every $\xi\in E'$,*

$$\sup_{\mu\in M}\int e^{\langle\xi,f(y)\rangle}\mu(dy)<\infty,$$

$$\sup_{x\in S}\int e^{\langle\xi,f(y)\rangle}\pi(x,dy)<\infty.$$

Let $T_\xi\colon B(S)\to B(S)$ be defined by

$$(T_\xi g)(x)=\int K_\xi(x,dy)\,g(y)$$

where

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

Then if $Y_n=\sum_{j=0}^{n-1} f(X_j)$ and $\xi\in E'$,

$$\limsup_n n^{-1}\log\sup_{\mu\in M}E_\mu\exp\langle\xi,Y_n\rangle\leqq\log r(T_\xi), \tag{4.1}$$

where $r(T_\xi)$ is the spectral radius of T_ξ.

Proof. First we need an expression for $E_\nu\exp\langle\xi,Y_n\rangle$. This is: for $n\geqq 2$,

$$E_\nu\exp\langle\xi,Y_n\rangle=\int\nu(dx)\,e^{\langle\xi,f(x)\rangle}K_\xi^{(n-1)}(x,S), \tag{4.2}$$

where $K_\xi^{(m)}$ is the m-fold composition power of the kernel K_ξ.

To prove (4.2) assume, inductively, that for a fixed $n\geqq 2$ and all initial distributions ν (4.2) holds.

Applying a suitable form of the Markov property (see, e.g., [12], p. 19), we have

$$\begin{aligned}E_\nu\exp\langle\xi,Y_{n+1}\rangle&=E_\nu E_\nu\{\exp\langle\xi,Y_{n+1}\rangle\,|\,\mathfrak{F}_1\}\\&=E_\nu\exp\langle\xi,f(X_0)\rangle\,E_\nu\{\exp\langle\xi,Y_n\circ\theta\rangle\,|\,\mathfrak{F}_1\}\\&=E_\nu\exp\langle\xi,f(X_0)\rangle\,E_{X_1}\exp\langle\xi,Y_n\rangle\end{aligned}$$

and by the inductive assumption

$$\begin{aligned}\dots &= E_\nu\{\exp\langle\xi, f(X_0)\rangle \exp\langle\xi, f(X_1)\rangle K_\xi^{(n-1)}(X_1, S)\}\\ &= \int \nu(dx)e^{\langle\xi, f(x)\rangle}\int e^{\langle\xi, f(y)\rangle}\pi(x, dy)K_\xi^{(n-1)}(y, S)\\ &= \int \nu(dx)\, e^{\langle\xi, f(x)\rangle} K_\xi^{(n)}(x, S).\end{aligned}$$

Since (4.2) is immediate for $n=2$, this proves (4.2). Next, by (4.2)

$$\begin{aligned}E_\mu \exp\langle\xi, Y_n\rangle &\leqq \sup_{x\in S} K_\xi^{(n-1)}(x, S)\int e^{\langle\xi, f(y)\rangle}\mu(dy)\\ &= \|T_\xi^{n-1}\| \int e^{\langle\xi, f(y)\rangle}\mu(dy),\end{aligned}$$

$$(\sup_{\mu\in M} E_\mu \exp\langle\xi, Y_n\rangle)^{1/n} \leqq \|T_\xi^{n-1}\|^{1/n}(\sup_{\mu\in M}\int e^{\langle\xi, f(y)\rangle}\mu(dy))^{1/n},$$

$$\begin{aligned}\limsup_n (\sup_{\mu\in M} E_\mu \exp\langle\xi, Y_n\rangle)^{1/n} &\leqq \lim_n \|T_\xi^{n-1}\|^{1/n}\\ &= r(T_\xi)\end{aligned}$$

by the spectral radius formula (see, e.g., [7], p. 567). This proves (4.1). □

For the statement of Theorem 4.2, let

$$\phi(\xi) = \log r(T_\xi) \qquad (\xi\in E');$$

also, as in Sect. 2, we write

$$\begin{aligned}\lambda(x) &= \sup_{\xi\in E'}[\langle\xi, x\rangle - \phi(\xi)] \qquad (x\in E),\\ \Lambda(A) &= \inf_{x\in A}\lambda(x) \qquad\qquad (A\subset E).\end{aligned}$$

Theorem 4.2. *Under the assumptions of Theorem* 4.1, *we have*

(a) *For every* $\sigma(E, E')$*-compact subset* F *of* E,

$$\limsup_n n^{-1}\log\sup_{\mu\in M} P_\mu\{n^{-1}Y_n\in F\} \leqq -\Lambda(F). \tag{4.3}$$

(b) *Let* E_0 *and* p *be as in Theorem* 3.1 *and assume* $f(S)\subset E_0$. *Assume also*

(i) *for every* $t>0$,

$$\sup_{\mu\in M}\int \exp\{tp(f(y))\}\,\mu(dy) < \infty,$$

$$\sup_{x\in S}\int \exp\{tp(f(y))\}\,\pi(x, dy) < \infty,$$

(ii) $\{f(\mu): \mu\in M\}$ *and* $\{\pi(x, f^{-1}(\cdot)): x\in S\}$ *are tight.*

Then (b_1) *for every closed set* $F\subset E$, (4.3) *holds,*

(b_2) *for every* $a\geqq 0$, $\{x\in E: \lambda(x)\leqq a\}$ *is compact.*

Proof. (a) follows at once from Theorems 2.1(i) and 4.1. In order to prove (b_1) it is enough, by Theorem 2.1(ii), to show that condition (2.3) is satisfied. By assumptions (i) and (ii) and Theorem 3.1, there exists a compact, convex, positively balanced set $K \subset E_0$, such that

$$\begin{aligned} b_1 &= \sup_{\mu\in M} \int \exp\{q_K(f(y))\}\, \mu(dy) < \infty, \\ b_2 &= \sup_{x\in S} \int \exp\{q_K(f(y))\}\, \pi(x,dy) < \infty. \end{aligned} \tag{4.4}$$

By the Markov property,

$$\begin{aligned} E_\mu \exp q_K(Y_n) &\leqq E_\mu \exp\{q_K(Y_{n-1}) + q_K(f(X_{n-1}))\} \\ &= E_\mu \exp\{q_K(Y_{n-1})\} \cdot \int \exp\{q_K(f(y))\}\, \pi(X_{n-2}, dy) \\ &\leqq (E_\mu \exp\{q_K(Y_{n-1})\}) \cdot b_2 \end{aligned}$$

and iterating we obtain, for all $\mu \in M$,

$$E_\mu \exp q_K(Y_n) \leqq b_1 b_2^{n-1} \leqq b^n,$$

where $b = \max\{b_1, b_2\}$. Therefore, for $\gamma > 0$, $\mu \in M$,

$$\begin{aligned} P_\mu\{n^{-1} Y_n \notin \gamma K\} &= P_\mu\{q_K(Y_n) > n\gamma\} \\ &\leqq e^{-n\gamma} E_\mu \exp q_K(Y_n) \\ &\leqq \exp\{-(\gamma - \log b)n\}. \end{aligned}$$

Given $a > 0$, choose $\gamma \geqq a + \log b$; then $K_a = \gamma K$ satisfies (2.3). This completes the proof of (b_1).

Let $a \geqq 0$ and let $F = \{x : \lambda(x) \leqq a\}$. We shall prove: there exists $0 < c < \infty$ such that if $x \in F$, then

$$\langle \xi, x\rangle \leqq c \quad \text{for all } \xi \in K^0, \tag{4.5}$$

where $K^0 = \{\xi \in E' : \langle \xi, x\rangle \leqq 1 \text{ for all } x \in K\}$, or equivalently,

$$c^{-1} x \in K^{00}.$$

Since $K^{00} = K$ by the bipolar theorem (see, e.g. [14], p. 126), it follows that

$$F \subset cK,$$

which proves (b_2).

In order to establish (4.5), let

$$H(x, A) = \int_A \exp(q_K(f(y)))\, \pi(x, dy),$$

$$(Tu)(x) = \int u(y) H(x, dy) \quad \text{for } u \in B(S);$$

then by (4.4), T: $B(S) \to B(S)$ is a bounded operator. Also, if $\xi \in K^0$,

$$\langle \xi, f(y)\rangle \leqq q_K(f(y))\, q_{K^0}(\xi) \leqq q_K(f(y)) \quad (y \in S)$$

and therefore

$$K_\xi^{(n)}(x, A) \leqq H^{(n)}(x, A) \quad (n \geqq 1).$$

It follows that for $\xi \in K^0$,

$$r(T_\xi) = \lim_n \|T_\xi^n\|^{1/n} \leqq \lim_n \|T^n\|^{1/n} = r(T).$$

Now for $x \in F$, for all $\xi \in E'$

$$\langle \xi, x \rangle \leqq a + \log r(T_\xi),$$

so for $\xi \in K^0$,

$$\langle \xi, x \rangle \leqq a + \log r(T) = c,$$

proving (4.5). □

Remarks. (a) It is not difficult to formulate and prove a result extending Theorem 4.2 to the case when $Y_n = \sum_{j=0}^{n-1} f(X_j, X_{j+1})$, where $f\colon S \times S \to E$. We omit the details.

(b) In the case of $\mathbb{R}^k$-valued functionals, sharper (that is, non-logarithmic) large deviation bounds for convex sets will be studied in a forthcoming paper by I. Iscoe, P. Ney and E. Nummelin.

We will apply Theorem 4.2 to two situations: the case of partial sums of independent, identically distributed random vectors and the case of occupation times of Markov chains. In both cases we obtain new proofs of the upper bound part of results of Donsker and Varadhan ([6], Theorems 4.4 and 5.3).

For the statement of Theorem 5.1, let

$$\hat{\mu}(\xi) = \int \exp\langle \xi, x \rangle \, \mu(dx) \qquad (\xi \in E'),$$

$$\lambda(x) = \sup_{\xi \in E'} [\langle \xi, x \rangle - \log \hat{\mu}(\xi)] \qquad (x \in E),$$

$$\Lambda(A) = \inf_{x \in A} \lambda(x) \qquad (A \subset E).$$

Theorem 4.3. *Let E, E_0 and p be as in Theorem 3.1. Let μ be a probability measure on E such that $\mu(E_0) = 1$ and*

$$\int \exp(tp)\, d\mu < \infty \qquad \text{for all } t > 0.$$

Let $\{X_j, j \geqq 1\}$ be a sequence of independent E-valued random vectors with $\mathscr{L}(X_j) = \mu$ for all j, and let $S_n = \sum_{j=1}^{n} X_j$. Then for every closed set $F \subset E$,

$$\limsup_n n^{-1} \log P\{n^{-1} S_n \in F\} \leqq -\Lambda(F). \tag{4.6}$$

Also, for every $a \geqq 0$ $\{x\colon \lambda(x) \leqq a\}$ is compact.

Proof. With obvious notational changes, the result follows from Theorem 4.2 by taking $S = E_0$, $f = \mathrm{Id}_S$, $M = \{\mu\}$, $\pi(x, \cdot) = \mu$ for all $x \in S$. Observe also that, trivially, $r(T_\xi) = \|T_\xi\| = \hat{\mu}(\xi)$. □

Remarks. (1) Theorem 4.3 is actually a slight generalization of Theorem 5.3 of [6], which follows by taking $(E, \|\cdot\|)$ to be a separable Banach space, $E_0 = E$ and $p = \|\cdot\|$.

(2) Again assume that E is a separable Banach space. Let $\{\mu_n\}$, μ be probability measures on E and assume that $\{\mu_n\}$ converges weakly to μ. For each $n\in\mathbb{N}$, let $\{X_j^{(n)}: 1\leqq j\leqq n\}$ be an independent system with $\mathscr{L}(X_j^{(n)})=\mu_n$, and let $S_n^{(n)}=\sum_{j=1}^{n} X_j^{(n)}$. Then (4.6) remains true with S_n replaced by $S_n^{(n)}$ under the integrability condition: for all $t>0$,

$$\sup_n \int \exp(t\|x\|)\,\mu_n(dx)<\infty.$$

This fact may be easily proved using Theorems 2.1 and 3.1. It is the upper bound part of a result proved by Chevet [4] for a sequence of Gaussian measures and by Bolthausen [3] for general probability measures.

For the formulation of the next theorem, which is essentially Theorem 4.4 of [6] we recall some definitions in [6]: for a Polish space S, $C_b(S)$ is the Banach space of real-valued bounded continuous functions defined on S, with the supremum norm;

$$\mathfrak{U}_1=\{u\in C_b(S): \inf_{x\in S} u(x)>0\};$$

π is a Feller transition probability: if $u\in C_b(S)$, then $\pi u\in C_b(S)$; for $\mu\in\mathfrak{M}_1^+(S)$, the space of probability measures on S (with the weak topology),

$$I(\mu)=\sup_{\mu\in\mathfrak{U}_1}\{\textstyle\int\log(u/\pi u)\,d\mu\},$$

$$I(A)=\inf_{\mu\in A} I(\mu) \qquad (A\subset\mathfrak{M}_1^+(S));$$

$\{L_n\}$ is the sequence of occupation times of the Markov chain $\{X_j\}$:

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

Theorem 4.4. *Let $\{X_j, j\geqq 0\}$ be a Markov chain with Polish state space S and Feller transition probability π. Let M be a relatively compact subset of $\mathfrak{M}_1^+(S)$. Then*

(a) *For every compact subset F of $\mathfrak{M}_1^+(S)$,*

$$\limsup_n n^{-1}\log\sup_{\mu\in M} P_\mu\{\omega: L_n(\omega,\cdot)\in F\}\leqq -I(F). \tag{4.7}$$

(b) *If, furthermore, $\{\pi(x,\cdot): x\in S\}$ is tight, then*

(b_1) *for every closed subset F of $\mathfrak{M}_1^+(S)$, (4.7) holds,*

(b_2) *for every $a\geqq 0$, $\{\mu\in\mathfrak{M}_1^+(S): I(\mu)\leqq a\}$ is compact.*

Proof. We apply Theorem 4.2. Let $f: S\to\mathfrak{M}_1^+(S)$ be defined by

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

then f is a continuous map and for all ω,

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

If $E=\mathfrak{M}(S)$, the space of finite signed measure on S endowed with the weak topology $\sigma(\mathfrak{M}(S), C_b(S))$, then $E'=C_b(S)$; if $\xi=\xi_g$ is the element of E' given by $g\in C_b(S)$, then

$$\langle \xi, \delta_y\rangle=\int g\,d\delta_y=g(y) \quad \text{for all } y\in S.$$

It is well known that if $E_0=\mathfrak{M}^+(S)$, the space of finite non-negative measures on S, then E_0 satisfies the assumption of Theorem 3.1. Moreover, if $p=\|\cdot\|_v$, the total variation norm, then p satisfies conditions (a) and (b) of Theorem 3.1 and $p(f(y))=\|\delta_y\|_v=1$ for all $y\in S$.

Assertions (a) and (b) above will follow from statements (a) and (b) of Theorem 4.2, respectively, if we prove: for all $\mu\in\mathfrak{M}_1^+(S)$,

$$I(\mu)=\lambda(\mu), \tag{4.8}$$

where λ is as in the statement of Theorem 4.2, taking into account that $E'=C_b(S)$; that is, λ is the conjugate convex function of

$$\phi(g)=\log r(T_g) \quad (g\in C_b(S)).$$

We first prove $I(\mu)\leqq\lambda(\mu)$. Let $u\in\mathfrak{U}_1$. Since π is a Feller transition probability, it follows that $g=\log(u/\pi u)\in C_b(S)$. We show next that $v=\pi u$ is an eigenvector of T_g associated to the eigenvalue 1. In fact, for all $x\in S$,

$$(T_g v)(x)=\int e^{g(y)}\,\pi(x,dy)\,v(y)=\int (u(y)/v(y))\,v(y)\,\pi(x,dy)=v(x). \tag{4.9}$$

We claim now that

$$r(T_g)=1. \tag{4.10}$$

For, since 1 is an eigenvalue of T_g, obviously $1\leqq r(T_g)$. To prove the reverse inequality, let $a=\inf_{x\in S} v(x)$ (necessarily positive). It follows from (4.9) that if $T=T_g$, then

$$v=T_n v\geqq a T^n 1,$$

$$\|v\|\geqq a\|T^n 1\|=\|T^n\|,$$

$$1=\lim_n(\|v\|/a)^{1/n}\geqq \lim_n \|T^n\|^{1/n}=r(T)$$

by the spectral radius formula (see, e.g., [7], p. 567). This proves the claim (4.10).

From (4.10) we have for $\mu\in\mathfrak{M}_1^+(S)$

$$\int\log(u/\pi u)\,d\mu=\int g\,d\mu-\log r(T_g)\leqq\lambda(\mu)$$

by the definition of λ. Since u may be arbitrarily chosen in $\mathfrak{U}_1$, it follows that

$$I(\mu)=\sup_{u\in\mathfrak{U}_1}\{\int\log(u/\pi u)\,d\mu\}\leqq\lambda(\mu).$$

Next we show that $\lambda(\mu)\leqq I(\mu)$. The following argument, which is more general than our original one, is taken from [15], p. 138. Given $g\in C_b(S)$, $a>\log r(T_g)$, introduce

$$u=\sum_{n=0}^{\infty} e^{-an}T_g^n 1.$$

Since $\|T_g^n 1\| \leqq \|T_g^n\|$, it follows from the spectral radius formula that the series converges uniformly and $u \in \mathfrak{U}_1$. Also,

$$\pi(e^g u) = e^a(u-1).$$

Let $v = e^g u$. Then $v \in \mathfrak{U}_1$ and

$$\int \log(v/\pi v)\, d\mu = \int \log\left(\frac{e^g u}{u-1}\right) - a$$
$$\geqq \int g\, d\mu - a.$$

It follows that

$$I(\mu) \geqq \int g\, d\mu - \log r(T_g),$$

and since $g \in C_b(S)$ is arbitrary this implies $I(\mu) \geqq \lambda(\mu)$. Thus (4.8) is proved. □

Remarks. (1) The tightness assumption in (b) of Theorem 4.4, though very close to Hypothesis (H^*) in [6], p. 415, appears to be slightly stronger.

(2) Statement (5.1) improves slightly the statement in Theorem 4.4 of [6], in which M is a (compact) set of point masses.

5. Sums of Exchangeable Random Vectors

Let B be a separable Banach space with Borel σ-algebra $\mathfrak{B}$ and let P be the distribution on $(B^{\mathbb{N}}, \mathfrak{B}^{\mathbb{N}})$ of an exchangeable sequence of B-valued random vectors. By the general version of de Finetti's theorem (see e.g. [5], p. 222, [11], p. 151), there exist a probability space $(M, \mathfrak{M}, \nu)$ and a transition probability μ on $M \times \mathfrak{B}$ such that

$$P = \int \tilde{\mu}_y\, \nu(dy),$$

where $\tilde{\mu}_y$ is the product measure on $(B^{\mathbb{N}}, \mathfrak{B}^{\mathbb{N}})$ will all marginals equal to μ_y.

For convenience, we shall take $(X_j)_{j \in \mathbb{N}}$ to be the sequence of coordinate functions on $(B^{\mathbb{N}}, \mathfrak{B}^{\mathbb{N}}, P)$. We write $S_n = \sum_{j=1}^{n} X_j$. As an application of Theorems 2.1 and 3.1, we obtain

Theorem 5.1. *Assume*

(1) *there exists a ν-null set $D \in \mathfrak{M}$ such that $\{\mu_y : y \in D^c\}$ is tight,*

(2) *for every $t > 0$,*

$$\left\| \int \exp(t\|x\|)\, \mu_{(\cdot)}(dx) \right\|_{L^\infty(\nu)} < \infty.$$

Then (a) *for every $\xi \in B'$,*

$$\lim_n n^{-1} \log E \exp\langle \xi, S_n \rangle = \phi(\xi),$$

where $\phi(\xi) = \log \left\| \int \exp(\langle \xi, x \rangle)\, \mu_{(\cdot)}(dx) \right\|_{L^\infty(\nu)}$;

(b) *for every closed set $F \subset B$,*

$$\limsup_n n^{-1} \log P\{n^{-1} S_n \in F\} \leqq -\Lambda(F),$$

where $\Lambda(F) = \inf_{x \in F} \lambda(x)$ and λ is the convex conjugate of ϕ.

(c) *for every* $a \geqq 0$, $\{x \in B: \lambda(x) \leqq a\}$ *is compact.*

Proof. (a)

$$\begin{aligned} E\exp\langle \xi, S_n\rangle &= \int \nu(dy) \int \exp(\langle \xi, S_n\rangle)\, d\tilde{\mu}_y \\ &= \int \nu(dy) \{\int \exp(\langle \xi, x\rangle)\, d\mu_y(x)\}^n \end{aligned}$$

and therefore $\lim_n (E\exp(\langle \xi, S_n\rangle))^{1/n} = \|\int \exp(\langle \xi, x\rangle)\, \mu_{(\cdot)}(dx)\|_{L^\infty(\nu)}$.

(b) We will show that condition (2.3) of Theorem 2.1 is satisfied. Then the statement follows from (a) and Theorem 2.1. By assumptions (1) and (2), there exists a ν-null set $E^c \in \mathfrak{M}$ such that $\{\mu_y: y \in E\}$ is tight and for every $t > 0$,

$$\sup_{y \in E} \int \exp(t\|x\|)\, \mu_y(dx) < \infty.$$

By Theorem 3.1, there exists a compact, convex, balanced set K such that

$$b = \sup_{y \in E} \int \exp(q_K)\, d\mu_y < \infty.$$

Then for $\gamma > 0$,

$$\begin{aligned} P\{S_n/n \notin \gamma K\} &= P\{q_K(S_n) > n\gamma\} \\ &\leqq e^{-n\gamma} E\exp q_K(S_n) \\ &= e^{-n\gamma} \int_E \nu(dy) \int \exp q_K(S_n)\, d\tilde{\mu}_y \\ &\leqq e^{-n\gamma} \int_E \nu(dy) (\int \exp(q_K)\, d\mu_y)^n \\ &\leqq e^{-n\gamma} b^n = e^{-n(\gamma - \log b)}. \end{aligned}$$

Thus for given $a > 0$, (2.3) is satisfied by choosing $\gamma \geqq a + \log b$ and $K_a = \gamma K$.

The proof of (c) is similar to that of (b_2) of Theorem 4.2. Let $\lambda(x) \leqq a$, $\xi \in K^0$. Then

$$\begin{aligned} \langle \xi, x\rangle &\leqq a + \phi(\xi) \\ &\leqq a + \log \|\int \exp(q_K(z))\, \mu_{(\cdot)}(dz)\|_{L^\infty(\nu)} \text{ since } \langle \xi, z\rangle \leqq q_K(z) \\ &\leqq a + \log b = c. \end{aligned}$$

Arguing as in the proof of (b_2) of Theorem 4.2, it follows that

$$\{x: \lambda(x) \leqq a\} \subset cK. \quad \square$$

Remark. Of course, the case of sums of independent identically distributed B-valued random vectors (Theorem 4.3) is also a corollary of Theorem 5.1.

Examples (1). Suppose that X_1 assumes only the values 0 and 1. Then by [10], p. 204, there exists a probability measure ν on $[0, 1]$ such that if $\mu_y(\{0\}) = 1 - y$, $\mu_y(\{1\}) = y$, then

$$P = \int \tilde{\mu}_y\, \nu(dy).$$

Theorem 5.1 applies and it is easily seen that for $\alpha \in \mathbb{R}$,

$$\phi(\alpha) = \begin{cases} \log(1 + S(e^\alpha - 1)) & \text{if } \alpha \geqq 0 \\ \log(1 + s(e^\alpha - 1)) & \text{if } \alpha < 0, \end{cases}$$

where $S = \sup(\text{support}(\nu))$, $s = \inf(\text{support}(\nu))$. The conjugate convex function λ may be easily computed in this case.

(2) Let μ be a centered Gaussian measure on B with covariance

$$\Phi(\xi, \eta) = \int \xi \eta \, d\mu \quad (\xi, \eta \in B').$$

For $\sigma \in \mathbb{R}^+$, let μ_σ be the centered Gaussian measure with covariance $\sigma \Phi$. Let ν be a probability measure on $\mathbb{R}^+$ with compact support and define on $(B^{\mathbb{N}}, \mathfrak{B}^{\mathbb{N}})$

$$P = \int \tilde{\mu}_\sigma \, \nu(d\sigma).$$

By well-known facts about Gaussian measures, assumptions (1) and (2) are satisfied and Theorem 5.1 applies. It is easily seen that

$$\phi(\xi) = (1/2) S \, \Phi(\xi, \xi),$$

where $S = \sup(\text{support}(\nu))$. The conjugate convex function is therefore the Cramér functional of the Gaussian measure μ_S (see [1]).

Acknowledgement. I wish to thank P. Ney for many stimulating discussions on large deviations and D. Griffeath for some useful seminar expositions on the subject.

References

1. Azencott, R.: Grandes déviations et applications. Lecture Notes in Mathematics **774**. Berlin-Heidelberg-New York: Springer 1980
2. Bahadur, R.R., Zabell, S.: Large deviations of the sample mean in general vector spaces. Ann. Probab. **7**, 587–621 (1979)
3. Bolthausen, E.: On the probability of large deviations in Banach spaces. Ann. Probab. **12**, 427–435 (1984)
4. Chevet, S.: Gaussian measures and large deviations. Lecture Notes in Mathematics **990**, 30–46. Berlin-Heidelberg-New York-Tokyo: Springer 1983
5. Chow, Y.S., Teicher, H.: Probability Theory. Berlin-Heidelberg-New York: Springer 1978
6. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time III. Commun. Pure Appl. Math. **29**, 389–461 (1976)
7. Dunford, N., Schwartz, J.: Linear operators I. New York: Interscience publishers 1967
8. Ellis, R.: Large deviations for a general class of dependent random vectors. Ann. Probab. **12**, 1–12 (1984)
9. Gärtner, J.: On large deviations from the invariant measure. Theor. Probab. Appl. **22**, 24–39 (1977)
10. Hall, P., Heyde, C.C.: Martingale limit theory and its application. New York: Academic Press 1980
11. Meyer, P.A.: Probability and potentials. Waltham, Mass.: Blaisdell 1966
12. Revuz, D.: Markov chains. Amsterdam: North Holland 1975
13. Rosinski, J.: Personal communication (1983)
14. Schaefer, H.H.: Topological vector spaces. New York: MacMillan 1966
15. Stroock, D.W.: An introduction to the theory of large deviations. Berlin-Heidelberg-New York-Tokyo: Springer 1984

Received February 10, 1984; in revised form November 10, 1984