The Annals of Probability
1984, Vol. 12, No. 1, 1–12

LARGE DEVIATIONS FOR A GENERAL CLASS OF RANDOM VECTORS

BY RICHARD S. ELLIS

University of Massachusetts

This paper proves large deviation theorems for a general class of random vectors taking values in $\mathbb{R}^d$ and in certain infinite dimensional spaces. The proofs are based on convexity methods. As an application, we give a new proof of the large deviation property of the empirical measures of finite state Markov chains (originally proved by M. Donsker and S. Varadhan). We also discuss a new notion of stochastic convergence, called exponential convergence, which is closely related to the large deviation results.

I. Introduction. This paper proves large deviation theorems for a general class of random vectors taking values in $\mathbb{R}^d$, $d \in \{1, 2, \cdots\}$. We also partially extend an upper large deviation estimate to a general infinite dimensional setting. These theorems complement some of the major large deviation results in the literature (e.g., Donsker-Varadhan, 1975, 1976, Bahadur-Zabell, 1979). The latter papers consider random vectors with a fairly general state space but a relatively simple dependence structure (i.i.d. or Markovian). The present paper handles any sequence of random vectors $\{Y_n\}$ for which the limit

$$c(t) = \lim_{n\to\infty}(1/a_n)\log E_n\{\exp\langle t,\ Y_n\rangle\}, \quad t \in \mathbb{R}^d,$$

exists for some sequence $a_n \to \infty$. This limit, which arises naturally in large deviation theory, is an analogue of the free energy in statistical mechanics. In fact, our results were inspired in part by the beautiful large deviation calculations for statistical mechanical random variables due to Lanford (1973). Applications of the present paper to statistical mechanics will appear in a forthcoming book (Ellis, 1984).

Our main analytic tool is the theory of convex functions and of the Legendre transform in particular. One purpose of this paper is to illustrate the power of these techniques in deriving large deviation results. We emphasize that proving these results for random vectors ($d \geq 1$) rather than just random variables ($d = 1$) is a nontrivial extension. As an application, we give an elementary proof of the large deviation property of the empirical measures of finite state Markov chains. This property was shown first by Donsker-Varadhan (1975) in much greater generality. Our new proof seems worthwhile since the proof of their general theorem was by necessity fairly involved.

Other papers have studied large deviations for general random vectors, but they differ in several respects from the present paper. Dacunha-Castelle (1979) treats only the case $d = 1$. (Although the proof in Dacunha-Castelle (1979) uses

Received November 1982; Revised May 1983.
AMS 1980 *subject classifications.* 60F10, 26A51.
Key words and phrases. Large deviation property, entropy function, exponential convergence.

the differentiability of the function $c(t)$ – see Section II below – the differentiability is not stated as a hypothesis.) Steinebach (1978) considers general $d \geq 1$, but he is interested in different large deviation estimates and his hypotheses are not the same as ours. Gärtner (1977) is closest to the present paper, but he does not prove the estimates for general closed and open subsets of $\mathbb{R}^d$ as we do.

Section II states the large deviation results, Section III contains the application to Markov chains, and Section IV discusses exponential convergence. Section V proves properties of the entropy function, then proves the theorems in Sections II and IV.

II. Statements of results.

IIa. *Finite dimensional case.* For simplicity, we consider random vectors indexed by the positive integers. With minor changes, our results go over to families of random vectors indexed by more general sets.

Let $\{\Omega_n, \mathscr{F}_n, P_n); n = 1, 2, \cdots\}$ be a sequence of probability spaces and $\mathscr{B}$ the Borel σ-algebra of $\mathbb{R}^d$. For each n, let Y_n be an $\mathscr{F}_n - \mathscr{B}$ measurable map of Ω_n into $\mathbb{R}^d$. Let $\mathscr{Y}$ be the sequence $\{Y_n; n = 1, 2, \cdots\}$. Given $t \in \mathbb{R}^d$, let

$$c_n(t) = c_n(\mathscr{Y}; t) \doteq (1/a_n)\log E_n\{\exp\langle t, Y_n\rangle\}, \tag{1}$$

where the $\{a_n\}$ are a fixed sequence of positive numbers tending to infinity, E_n denotes expectation with respect to P_n, and $\langle -, -\rangle$ is the Euclidean inner product on $\mathbb{R}^d$.

HYPOTHESIS II.1. *$c(t) = c(\mathscr{Y}; t) \doteq \lim_{n\to\infty} c_n(t)$ exists for all $t \in \mathbb{R}^d$, where we allow $+\infty$ both as a limit value and as an element in the sequence $\{c_n(t)\}$. (Define $c(t) = \infty$ if $c_n(t) = \infty$ for all $n \geq n_0$ (n_0 depending on t)). Let $\mathscr{D}(c) \doteq \{t \in \mathbb{R}^d : c(t) < \infty\}$. $\mathscr{D}(c)$ has non-empty interior containing the point $t \doteq 0$, and c is a closed convex function on $\mathbb{R}^d$.*

$\mathscr{D}(c)$ is always non-empty since $c(0) = 0 < \infty$. The convexity of c follows from the convexity of each c_n. By definition, c is closed if for each real α, the set $\{t \in \mathbb{R}^d : c(t) \leq \alpha\}$ is closed in $\mathbb{R}^d$. This is equivalent to c being lower semicontinuous. If c is differentiable on the interior of $\mathscr{D}(c)$, int $\mathscr{D}(c)$, then we call c *steep* if $\|\operatorname{grad} c(x_n)\| \to \infty$ for any sequence $\{x_n\} \subseteq \operatorname{int} \mathscr{D}(c)$ which tends to a boundary point of $\mathscr{D}(c)$. If c is closed and $\mathscr{D}(c)$ is an open set, then c is steep (Barndorff-Nielsen, 1978; Corollary 5.3). For the theory of convex functions on $\mathbb{R}^d$, Rockafellar (1970) is the standard reference.

Let Q_n be the distribution of $a_n^{-1} Y_n$ and let $I = I(\mathscr{Y}; \cdot)$ be the Legendre/Fenchel transform of c:

$$I(z) \doteq \sup_{t\in\mathbb{R}^d}\{\langle t, z\rangle - c(t)\}, \quad z \in \mathbb{R}^d. \tag{2}$$

Given a subset A of $\mathbb{R}^d$, define $I(A) \doteq \inf\{I(z) : z \in A\}$. Here is our first large deviation result.

THEOREM II.2. *We assume Hypothesis* II.1.

(a) *For any closed subset K of* $\mathbb{R}^d$,

$$\limsup_{n\to\infty}(1/a_n)\log Q_n\{K\} \le -I(K). \tag{3}$$

(b) *If c is differentiable on all of* int $\mathscr{D}(c)$ *and is steep, then for any open subset G of* $\mathbb{R}^d$

$$\liminf_{n\to\infty}(1/a_n)\log Q_n\{G\} \ge -I(G). \tag{4}$$

If (3) and (4) hold for all closed K and all open G, respectively, then we say that $\{Q_n\}$ has a *large deviation property with entropy function I.* A special case of this theorem is where Y_n is the nth partial sum of i.i.d. random vectors $X_1, X_2, \cdots$ for which $E\{\exp\langle t, X_1\rangle\} < \infty$ for all $t \in \mathbb{R}^d$. Then Hypothesis II.1 is valid with $c(t) = \log E\{\exp\langle t, X_1\rangle\}$. This case is well-known.

In general, if c is not differentiable on all int $\mathscr{D}(c)$, then (4) is not valid for certain open G. Here is a simple example. Let Y_n have distribution $(\delta_n + \delta_{-n})/2$. Then (with $a_n = n$) we have $c(t) = |t|$, $t \in \mathbb{R}$, which is closed and steep, and $I(z) = 0$ if $|z| \le 1$, ∞ if $|z| > 1$. If G is an open subset of the interval $(-1, 1)$, then the left-hand side of (4) equals $-\infty$ while the right-hand side equals 0. Thus (4) fails for these G. On the other hand, it is easy to see that the distributions $\{Q_n\}$ of $\{Y_n\}$ have a large deviation property with entropy function

$$I(z) = \begin{cases} 0 & \text{for} \quad z = 1, \quad z = -1 \\ \infty & \text{for} \quad z \notin \{1, -1\} \end{cases}$$

This entropy function is not convex.

IIb. *Infinite dimensional case.* We extend the upper bound (3) for compact sets. We can also partially extend the lower bound (4) (see the paragraph after (30)), and Theorem V.1. But we omit these.

Let $\mathscr{Z}$ be a real vector space of points z and τ a given topology on $\mathscr{Z}$. We assume that under τ, $\mathscr{Z}$ is a locally convex Hausdorff topological space (l.c.H.t.s.). Let $\mathscr{T}$ be the topological dual of $\mathscr{Z}$ ($\mathscr{T} \doteq \mathscr{Z}^*$) and for $t \in \mathscr{T}$ and $z \in \mathscr{Z}$ let $\langle t, z\rangle$ denote the value at z of the continuous linear functional t. We consider $\mathscr{T}$ with the topology of weak convergence over $\mathscr{Z}$. Then $\mathscr{T}$ is also an l.c.H.t.s. Let $\{(\Omega_n, \mathscr{F}_n, P_n); n = 1, 2, \cdots\}$ be a sequence of probability spaces and $\mathscr{B}$ the σ-algebra of Borel subsets of $\mathscr{Z}$. For each n, let Y_n be an $\mathscr{F}_n - \mathscr{B}$ measurable map of Ω_n into $\mathscr{Z}$. Given $t \in \mathscr{T}$ we define $c_n(t) = c_n(\mathscr{Y}; t)$ as in (1), where $\{a_n\}$ and E_n are as above.

HYPOTHESIS II.3. *$c(t) \doteq \lim_{n\to\infty} c_n(t)$ exists for all $t \in \mathscr{T}$, where we allow $+\infty$ both as a limit value and as an element in the sequence $\{c_n(t)\}$. (Define $c(t) = \infty$ if $c_n(t) = \infty$ for all $n \ge n_0$ (n_0 depending on t)).*

Let Q_n be the distribution of $a_n^{-1}Y_n$ and let $I = I(\mathscr{Y}; \cdot)$ be the Legendre/

Fenchel transform of c:

$$I(z) \doteq \sup_{t \in \mathscr{T}} \{\langle t, z\rangle - c(t)\}, \quad z \in \mathscr{Z}. \tag{5}$$

I is well-defined (see Ekeland-Temam, 1976, Chapter I).

THEOREM II.4. *We assume Hypothesis* II.3. *Then for any compact subset K of $\mathscr{Z}$*

$$\limsup_{n\to\infty}(1/a_n)\log Q_n\{K\} \leq -I(K). \tag{6}$$

III. Application. We prove a special case of a result in Donsker-Varadhan (1975). Let $\{X_1, X_2, \cdots\}$ be a Markov chain with state space Γ consisting of d distinct real numbers $\{x_1, \cdots, x_d\}$ ($d \geq 2$). The chain starts at a fixed point $x \in \Gamma$. Let P_x be the probability measure for the chain and $\pi(i, j) \doteq \pi(\{x_i\}, \{x_j\})$ its transition probabilities. We assume that the matrix $\pi \doteq \{\pi(i, j)\}$ is irreducible and aperiodic. The chain then has a unique invariant measure ρ. For each subset A of Γ we define the empirical measure of A by the formula $L_n(\omega)\{A\} \doteq (1/n)\sum_{j=1}^n \delta_{X_j(\omega)}\{A\}$. By the ergodic theorem

$$L_n \Rightarrow_w \rho \quad \text{a.s.} \quad (P_x). \tag{7}$$

Let $\mathscr{M}$ be the set of probability measures on Γ. L_n takes values in $\mathscr{M}$. Any $\nu \in \mathscr{M}$ can be represented as $(\nu_1, \cdots, \nu_d)$, where $\nu_i \doteq \nu(\{x_i\}) \geq 0$ and $\sum \nu_i = 1$. We identify $\mathscr{M}$ with this subset of $\mathbb{R}^d$. Part (b) of the next theorem is new.

THEOREM III.1. (a) *Let $\{Q_n\}$ be the distributions of $\{L_n\}$ in $\mathscr{M}$. Then $\{Q_n\}$ has a large deviation property with entropy function*

$$I_\pi(\nu) \doteq -\inf_{u>0} \sum_{i=1}^d \nu_i \log\frac{(\pi u)_i}{u_i}, \quad \nu \in \mathscr{M}, \tag{8}$$

where $u \doteq (u_1, \cdots, u_d)$, each $u_i > 0$, and $(\pi u)_i \doteq \sum_{j=1}^d \pi(i, j)u_j$; $I_\pi(\nu) = \infty$ for $\nu \in \mathbb{R}^d\backslash\mathscr{M}$.

(b) *More generally, let $\{\pi^{(n)}\}$ be a sequence of irreducible, aperiodic stochastic matrices which tend to an irreducible, aperiodic stochastic matrix π. Then the conclusion of* (a) *holds for the distributions of the empirical measures based on $\{\pi^{(n)}\}$.*

PROOF. (a) Large deviations for $\{L_n\}$ are equivalent to those for the random vectors $\{n^{-1}Y_n\}$, where Y_n has ith component $Y_{n,i} \doteq \sum_{j=1}^n \delta_{X_j}\{x_i\}$. Given $t \in \mathbb{R}^d$, define $t(x_i) \doteq t_i$ for $x_i \in \Gamma$. Then

$$c_n(t) = (1/n)\log E_x\{\exp(\textstyle\sum_{j=1}^n t(X_j))\}, \tag{9}$$

where E_x denotes expectation with respect to P_x. Let $B = B(\pi, t)$ denote the matrix $\{\exp(t_i)\cdot\pi(i, j)\}$. A short calculation shows that

$$c_n(t) = (1/n)\log \textstyle\sum_{i_1, i_n=1}^r \delta_x(x_{i_1})B^{n-1}(i_i, i_n)\exp(t_{i_n}). \tag{10}$$

Let B be an irreducible, aperiodic, non-negative matrix (such a matrix is also called *primitive.* (Berman/Plemmons, 1979)). Its largest eigenvalue in absolute value, $\lambda(B)$, is simple and for each i, j, $[B^n(i, j)]^{1/n} \to \lambda(B)$ (Spitzer, 1971; Theorem 1.9, 1.11). Hence for each $t \in \mathbb{R}^d$

$$\lim_{n\to\infty} c_n(t) = c(t) \doteq \log \lambda(B(\pi, t)). \tag{11}$$

$c(t)$ is closed, differentiable, and steep, and so the conclusions of Theorem II.2 hold with

$$I(\nu) \doteq \sup_{t\in\mathbb{R}^d}\{\langle t, \nu\rangle - \log \lambda(B(\pi, t))\}. \tag{12}$$

The next lemma allows us to reduce (12) to (8).

LEMMA III.2. *Let $\mathscr{N}^+$ denote the set of all $t \in \mathbb{R}^d$ which are of the form $t_i = \log[u_i/(\pi u)_i]$ for some $u > 0$ in $\mathbb{R}^d$. Then $\lambda(B(\pi, t)) = 1$ if and only if $t \in \mathscr{N}^+$. Furthermore, any $t \in \mathbb{R}^r$ can be written as $\bar{t} + c\mathbf{1}$ for some $\bar{t} \in \mathscr{N}^+$ and real c, where $\mathbf{1}$ denotes the constant vector* $(1, \cdots, 1)$.

In (12) we replace t by $\bar{t} + c\mathbf{1}$ as in Lemma III.2. For $\nu \in \mathscr{M}$

$$\langle \nu, \bar{t} + c\mathbf{1}\rangle - \log \lambda(B(\pi, \bar{t} + c\mathbf{1})) = \langle \nu, \bar{t}\rangle - \log \lambda(B(\pi, \bar{t})), \tag{13}$$

and since $\bar{t}$ is in $\mathscr{N}^+$, the last term is zero. Since $\bar{t}_i = \log(u_i/(\pi u)_i)$ for some $u > 0$, (12) becomes

$$I(\nu) = \sup_{u>0} \sum_{j=1}^{d} \nu_j \log \frac{u_j}{(\pi u)_j}. \tag{14}$$

This is exactly (8). We must prove that $I(\nu)$ in (12) is ∞ for $\nu \in \mathbb{R}^d\backslash\mathscr{M}$. Since L_n takes values in $\mathscr{M}$, this follows from the lower large deviation bound (4).

(b) For each t, $B(\pi^{(n)}, t) \to B(\pi, t)$ and so $\lambda(B(\pi^{(n)}, t)) \to \lambda(B(\pi, t))$ (Kato, 1966, Theorem II.5.1). This is the same function $c(t)$ as in part (a).□

PROOF OF LEMMA III.2. We write B for $B(\pi, t)$ and λ for $\lambda(B(\pi, t))$. Say $\lambda = 1$. Then B has a corresponding right eigenvector $u > 0$, and $\exp(t_i) \cdot (\pi u)_i = u_i$ for each i. This implies that $t \in \mathscr{N}^+$. If $t \in \mathscr{N}^+$, then $t_i = \log(u_i/(\pi u)_i)$ for some $u > 0$ and u is an eigenvector of B corresponding to the eigenvalue 1. If λ exceeded 1, then by the proof of Theorem 1.11 in Spitzer (1971) both $\lambda^{-n}B^n(i, j)$ and $B^n(i, j)$ would have positive limits for all i and j. This is impossible, and so $\lambda = 1$. To prove the last assertion, given $t \in \mathbb{R}^d$ define $c \doteq \log \lambda(B(\pi, t))$ and $\bar{t} \doteq t - c$. This $\bar{t}$ is in $\mathscr{N}^+$. □

IV. Exponential convergence. We discuss a new notion of stochastic convergence which is related to the large deviation results. Let $\mathscr{Y} \doteq \{Y_n\}$ and $\{a_n\}$ be as in Section IIa. Given $\gamma \in \mathbb{R}^d$, we say that $\{a_n^{-1}Y_n\}$ *tends to* γ *exponentially*, and write $a_n^{-1}Y_n \to_{\exp} \gamma$, if for any sufficiently small $\varepsilon > 0$ there exists $M = M(\varepsilon) > 0$ such that

$$P_n\{\| a_n^{-1}Y_n - \gamma \| \geq \varepsilon\} \leq \exp(-a_nM) \quad \text{for all sufficiently large} \quad n. \tag{15}$$

If $a_n^{-1} Y_n \to_{\exp} \gamma$, then we say that $\{a_n^{-1} Y_n\}$ *satisfies the exponential law of large numbers*. If the $\{Y_n\}$ are all defined on the same probability space and $\sum_{n\geq 1} \exp(-a_n M) < \infty$ for all $M > 0$, then exponential convergence implies almost sure convergence. This is a direct consequence of the Borel-Cantelli Lemma.

Our first result shows that $\{a_n^{-1} Y_n\}$ satisfies the exponential law of large numbers if and only if $c(\mathscr{Y}; t)$ is differentiable at $t \doteq 0$. The equivalence between (b) and (c) is an aspect of Legendre/Fenchel duality, which will be explored further in the next section.

THEOREM IV.1. *We assume Hypothesis* II.1. *The following three statements are equivalent.*

(a) $a_n^{-1} Y_n \to_{\exp} \gamma$ *for some* $\gamma \in \mathbb{R}^d$.
(b) $c = c(\mathscr{Y}; \cdot)$ *is differentiable at* $t \doteq 0$ *and* grad $c(0) = \gamma$.
(c) $I = I(\mathscr{Y}; \cdot)$ *achieves its global minimum at the unique point* γ.

In statistical mechanics, the parameters $\{a_n\}$ represent the numbers of particles in a sequence of physical systems indexed by $\{n\}$. These particles may assume different configurations (e.g., molecules in a gas, spin configurations in a magnet). The $\{Y_n\}$ represent configuration-dependent quantities (e.g., total energy, total spin) which are proportional to $\{a_n\}$. The inequality (15) states that in the limit $n \to \infty$, almost all configurations have the same value γ of $a_n^{-1} Y_n$. This gives a scheme for deriving the stable thermodynamics of macroscopic systems from the chaotic behavior of the individual particles which constitute the system. A nice application of Theorem IV.1 is to convergence properties of the total spin in Ising and related models of ferromagnetism. These properties include the case where the total spin fails to converge exponentially to a constant, this being equivalent to the ferromagnetic phase transition (called spontaneous magnetization). The monograph Ellis (1984) will treat these and other applications of Theorem IV.1.

V. Properties of the entropy function and proofs of theorems in Sections II and IV. The following theorem will be used to prove the theorems in Sections II and IV. In parts (c), (d), and (g), ∂c and ∂I denote the subdifferentials of c and I.

THEOREM V.1. *We assume Hypothesis* II.1.

(a) $I = I(\mathscr{Y}; \cdot)$ *is a closed convex function on* $\mathbb{R}^d$.
(b) *For all* $t \in \mathscr{D}(c)$ *and* $z \in \mathscr{D}(I)$, $\langle t, z\rangle \leq c(t) + I(z)$.
(c) $\langle t, z\rangle = c(t) + I(z)$ *if and only if* $z \in \partial c(t)$.
(d) $z \in \partial c(t)$ *if and only if* $t \in \partial I(z)$.
(e) $c(t) = \sup_{z\in\mathbb{R}^d}\{\langle t, z\rangle - I(z)\}$, $t \in \mathbb{R}^d$.
(f) *For each* α *real, the set* $L_\alpha \doteq \{z : I(z) \leq \alpha\}$ *is a closed, bounded, convex subset of* $\mathbb{R}^d$.
(g) $\inf\{I(z) : z \in \mathbb{R}^d\}$ *is* 0, *and* $I(z_0) = 0$ *if and only if* $z_0 \in \partial c(0)$, *which is a nonempty, closed, bounded, convex subset of* $\mathbb{R}^d$.

PROOF. (a)–(e) Rockafellar (1970; Theorems 12.2 and 23.5).

(f) By Hypothesis II.1 there exists $\varepsilon > 0$ and a ball $B_{2\varepsilon}$ of radius 2ε and center 0 which is contained in int $\mathscr{D}(c)$. For any $z \in L_\alpha$.

$$\sup_{t \in B_\varepsilon} | \langle t, z \rangle | = \varepsilon \| z \| \le \alpha + \sup_{t \in B_\varepsilon} | c(t) | < \infty, \tag{16}$$

and so L_α is bounded. The other properties are obvious.

(g) By the definition of subdifferential $I(z)$ achieves its global minimum at $z_0 \Leftrightarrow 0 \in \partial I(z_0) \Leftrightarrow z_0 \in \partial c(0)$ by (d). By (c), $I(z_0) = 0$ for any $z_0 \in \partial c(0)$. The properties of $\partial c(0)$ follow from Hypothesis II.1, Rockafellar (1970; Theorem 23.4), and the fact that $\partial c(0)$ equals the set L_0 in (f). □

PROOF OF THEOREM II.2. To prove the upper bound (3), we reduce the case of an arbitrary closed set K to the case of closed half spaces. The next lemma is the key. It relies strongly upon Lemma 1.1 in Gärtner (1977).

LEMMA V.2. *Given $t \in \mathscr{D}(c)$ and α real, define the closed half space*

$$H_+(t, \alpha) \doteq \{z \in \mathbb{R}^d : \langle t, z \rangle - c(t) \ge \alpha\}. \tag{17}$$

Let K be a closed subspace of $\mathbb{R}^d$. If $0 < I(K) < \infty$, then for any $\varepsilon \in (0, I(K))$ there exists finitely many points $t_1, \cdots, t_r$ in $\mathscr{D}(c)$ such that

$$K \subseteq \cup_{i=1}^r H_+(t_i, I(K) - \varepsilon). \tag{18}$$

If $I(K) = \infty$, then for any $R > 0$ there exist finitely many points $t_1, \cdots, t_r$ in $\mathscr{D}(c)$ such that

$$K \subseteq \cup_{i=1}^r H_+(t_i, R). \tag{19}$$

We prove the lemma in a moment. If $I(K) = 0$, then (3) is clear. Say $0 < I(K) < \infty$. By (18) with $\alpha \doteq I(K) - \varepsilon > 0$, Chebyshev's inequality implies

$$\begin{aligned} P_n\left\{\frac{Y_n}{a_n} \in K\right\} &\le \textstyle\sum_{i=1}^r P_n\left\{\frac{Y_n}{a_n} \in H_+(t_i, \alpha)\right\} \\ &= \textstyle\sum_{i=1}^r P_n\{\langle t_i, Y_n \rangle \ge a_n(c(t_i) + \alpha)\} \\ &\le \textstyle\sum_{i=1}^r \exp\{a_n(c_n(t_i) - c(t_i) - \alpha)\}. \end{aligned} \tag{20}$$

This yields (3) since $c_n(t_i) \to c(t_i) < \infty$. If $I(K) = \infty$, then (19) yields (3).

PROOF OF LEMMA V.2. We prove only the case $0 < I(K) < \infty$; the case $I(K) = \infty$ is handled similarly. Let $A \doteq \{z \in \mathbb{R}^d : I(z) \le I(K) - \varepsilon\}$. By Theorem V.1 (f), this set is compact. Let S be a closed ball containing A such that the boundary of S (bd S) and A are disjoint. Define $U \doteq (K \cap S) \cup \text{bd } S$. We shall find finitely many points $t_1, \cdots, t_r$ in $\mathscr{D}(c)$ such that

$$U \subseteq \cup_{i=1}^r H_+(t_i, I(K) - \varepsilon). \tag{21}$$

Afterwards we prove that (21) implies (18).

We write $H_+(t)$ for the closed half space $H_+(t, I(K) - \varepsilon)$ and $H_-(t)$ for the opposite closed half space. By the definition of $I(z)$, $A = \cap_{t \in \mathscr{D}(c)} H_-(t)$. Hence $A^c = \cup_{t \in \mathscr{D}(c)} \text{int } H_+(t)$. Since $U \subseteq A^c$, for each $z \in U$ there exists an open neighborhood $N(z)$ of z and a point $t \in \mathscr{D}(c)$ such that $N(z) \subseteq H_+(t)$. The compactness of U implies (21).

We now prove (18). Since $K \cap S^c \subseteq S^c$, it suffices to prove that any $x \in S^c$ belongs to $\cup_{i=1}^r H_{+,i}$, where $H_{+,i} \doteq H_+(t_i)$. Say there exists an $x \in S^c$ for which this fails. Then $x \in \cap_{i=1}^r \text{int } H_{-,i}$. Pick any $\theta \in A$, which is a subset of $\cap_{i=1}^r H_{-,i}$. Since $\cap_{i=1}^r \text{int } H_{-,i}$ is convex, the interval

$$(\theta, x] \doteq \{z\colon z = \lambda\theta + (1 - \lambda)x, 0 \leq \lambda < 1\} \tag{22}$$

belongs to $\cap_{i=1}^r \text{int } H_{-,i}$. By choice of S, $(\theta, x]$ intersects bd S at some point b. Thus $b \in \cap_{i=1}^r \text{int } H_{-,i}$, which contradicts (21). □

We now prove the lower bound (4), first under the assumption that the relative interior of $\mathscr{D}(I)$ (ri $\mathscr{D}(I)$) is nonempty. The case where $\mathscr{D}(I)$ is a single point will be considered later. As in the proof of Lemma 1.2 in Gärtner (1977), we need the following key fact.

LEMMA V.3. *We assume that Hypothesis* II.1 *holds and that c is differentiable on all of* int $\mathscr{D}(c)$ *and is steep. Let* $\mathscr{F}$ *be the image of* int $\mathscr{D}(c)$ *under* grad c. *Then* ri $\mathscr{D}(I) \subseteq \mathscr{F} \subseteq \mathscr{D}(I)$.

PROOF. Rockafellar (1970; Corollary 26.4.1). □

If G is an open set, then $I(G) = I(G \cap \mathscr{D}(I))$, where we define $I(\phi) = \infty$. By the continuity property in Rockafellar (1970; Corollary 7.5.1), $I(G) = I(G \cap \text{ri } \mathscr{D}(I))$. Hence it suffices to prove

$$\liminf_{n\to\infty}(1/a_n)\log Q_n\{G\} \geq -I(G \cap \text{ri } \mathscr{D}(I)), \tag{23}$$

where we assume that $G \cap \text{ri } \mathscr{D}(I)$ is non-empty. For $t \in \mathscr{D}(c_n)$, let

$$dQ_{n,t}(x) \doteq \frac{\exp(a_n\langle t, x\rangle)\, dQ_n(x)}{\exp(a_n c_n(t))}. \tag{24}$$

Given $z \in G \cap \text{ri } \mathscr{D}(I)$, we pick $\varepsilon > 0$ such that $B(z, \varepsilon)$, the open ball of radius ε centered at z, is contained in G. By Lemma V.3, there exists a point $t = t(z) \in$ int $\mathscr{D}(c)$ such that grad $c(t) = z$. For all sufficiently large n, $c_n(t) < \infty$. With this t, $\langle t, z\rangle - c(t) = I(z)$ (Theorem V.1(c)). For $x \in B(z, \varepsilon)$, $-\langle t, x\rangle \geq -\langle t, z\rangle - \varepsilon \| t \|$. Thus for all sufficiently large n

$$Q_n\{G\} \geq Q_n\{B(z, \varepsilon)\} = \exp(a_n c_n(t)) \int_{B(z,\varepsilon)} \exp(-a_n\langle t, x\rangle)\, dQ_{n,t}(x), \tag{25}$$

$$\begin{aligned} &\liminf_{n\to\infty}(1/a_n)\log Q_n\{G\} \\ &\qquad \geq c(t) - \langle t, z\rangle - \varepsilon \| t \| + \liminf_{n\to\infty}(1/a_n)\log Q_{n,t}\{B(z, \varepsilon)\}. \end{aligned} \tag{26}$$

We prove that

$$\lim_{n\to\infty} Q_{n,t}\{B(z,\varepsilon)\} = 1. \tag{27}$$

Taking $\varepsilon \to 0$ in (26), we conclude that

$$\liminf_{n\to\infty}(1/a_n)\log Q_n\{G\} \geq -I(z). \tag{28}$$

Taking the supremum over $z \in G \cap \text{ri } \mathscr{D}(I)$ yields (23).

Let $a_n^{-1}\tilde{Y}_n$ have distribution $dQ_{n,t(z)}$ and let $\tilde{\mathscr{Y}} \doteq \{\tilde{Y}_n\}$. For all $s \in \mathbb{R}^d$ with $\| s \|$ sufficiently small

$$\begin{aligned} c(\tilde{\mathscr{Y}}; s) &\doteq \lim_{n\to\infty} \frac{1}{a_n} \log \int_{\mathbb{R}^d} \exp(a_n\langle s, x\rangle)\, dQ_{n,t(z)}(x) \\ &= c(\mathscr{Y}; t(z) + s) - c(\mathscr{Y}; t(z)). \end{aligned} \tag{29}$$

Hence $c(\tilde{\mathscr{Y}}; \cdot)$ is differentiable at 0, and (27) follows from (b) $\Rightarrow$ (a) in Theorem IV.1. (The proof of the latter relies only on the upper bound (3), which has already been proved. It does not depend on the lower bound (4), which is now being proved.)

We now consider the case where ri $\mathscr{D}(I)$ is empty and $\mathscr{D}(I)$ is a single point $\{z_0\}$. Then $I(z_0) = 0$, and $c(t) = \langle t, z_0\rangle$ for all $t \in \mathbb{R}^d$. For all t, $\langle t, z_0\rangle - c(t) = I(z_0)$. By the same steps used to prove (28), one shows that if G is any open set containing z_0, then

$$\liminf_{n\to\infty}(1/a_n)\log Q_n\{G\} \geq -I(z_0). \tag{30}$$

This yields (4) and completes the proof of Theorem II.2. □

Assume that c is differentiable on only a subset Δ of int $\mathscr{D}(c)$ and let

$$\mathscr{F} \doteq \{z : z = \text{grad } c(t) \text{ for some } t \in \Delta\}. \tag{31}$$

Then the lower bound (4) holds for any open set G in $\mathscr{F}$.

Proof of Theorem II.4. We need the analogue of Lemma V.2 for K a compact subset of $\mathscr{Z}$. Again we prove only the case $0 < I(K) < \infty$. The set $A \doteq \{z \in \mathscr{Z} : I(z) \leq I(K) - \varepsilon\}$ is closed and is disjoint from K. For $t \in \mathscr{D}(c)$, let $H_+(t)$ denote the closed half-space $H_+(t, I(K) - \varepsilon)$, $\varepsilon \in (0, I(K))$, defined as in (17). Since

$$K \subseteq A^c = \cup_{t\in\mathscr{D}(c)} \text{int } H_+(t), \tag{32}$$

for each $z \in \mathscr{Z}$ there exists an open neighborhood $N(z)$ of z and a point t in $\mathscr{D}(c)$ such that $N(z) \subseteq H_+(t)$. The compactness of K implies (18). The latter yields the upper bound (6) exactly as in the finite dimensional case. □

Proof of Theorem IV.1.

(b) $\Leftrightarrow$ (c) Theorem V.1(g).

(c) $\Rightarrow$ (a) Given $\varepsilon > 0$, let K be the closed set $\{z \in \mathbb{R}^d : \| z - \gamma \| \geq \varepsilon\}$. By the upper bound (3), for all n sufficiently large

$$P_n\left\{\frac{Y_n}{a_n} \in K\right\} \le \exp\left(-\frac{a_n}{2} I(K)\right). \tag{33}$$

Hence the exponential bound follows once we prove that $I(K) > 0$. If $I(K) = 0$, then there exists a sequence $\{z_n\}$ in K such that $I(z_n) \to 0$. By Theorem V.1(f), we may assume that $z_n \to \tilde{z}$ for some $\tilde{z}$. Since K is closed, $\tilde{z}$ is in K. Since I is lower semicontinuous (closed) and non-negative, $I(\tilde{z}) = 0$. We obtain a contradiction since K does not contain the point γ and by hypothesis $I(z)$ achieves its minimum at the unique point $z = \gamma$.

(a) $\Rightarrow$ (b) Let s be an arbitrary unit vector in $\mathbb{R}^d$. Define $D_s^+c(0)$ and $D_s^-c(0)$ to be the right-hand and left-hand derivatives at 0 of the function $\mu \in \mathbb{R} \to c(\mu s)$. By Rockafellar (1970; Theorem 25.2) it suffices to prove

$$D_s^+c(0) = D_s^-c(0) = \langle s, \gamma\rangle. \tag{34}$$

Let $J_s \doteq \{\bar{\mu} > 0 : 2\mu s \in \operatorname{int} \mathscr{D}(c) \text{ for all } |\mu| \le \bar{\mu}\}$. J_s is non-empty since $0 \in \operatorname{int} \mathscr{D}(c)$. For any $\bar{\mu} \in J_s$, $0 < |\mu| \le \bar{\mu}$, and all sufficiently large n, $c_n(\mu s)$ is finite, and

$$\frac{c_n(\mu s)}{\mu} - \langle s, \gamma\rangle = \frac{1}{a_n\mu} \log E_n\{\exp[a_n\mu\langle s, V_n\rangle]\}, \tag{35}$$

where $V_n \doteq a_n^{-1}Y_n - \gamma$. Given $\varepsilon > 0$, we divide the latter expectation into two parts: the first over the set where $\| V_n \| < \varepsilon$ and the second over the set where $\| V_n \| \ge \varepsilon$. The first is bounded above by $\exp(a_n |\mu| \varepsilon)$. For the second, we have

$$\begin{aligned} &E_n\{\exp[a_n\mu\langle s, V_n\rangle]; \| V_n \| \ge \varepsilon\} \\ &\qquad \le \exp(-a_n\mu\langle s, \gamma\rangle)[E_n\{\exp[2\mu\langle s, Y_n\rangle]\}]^{1/2}[P_n\{\| V_n \| \ge \varepsilon\}]^{1/2}. \end{aligned} \tag{36}$$

By part (a) of Theorem IV.1, $V_n \to_{\exp} 0$. Putting these facts together, we conclude that there exists $\bar{\varepsilon} > 0$ and for all $0 < \varepsilon \le \bar{\varepsilon}$ there exists $M(\varepsilon) > 0$ such that for all $0 < \mu \le \bar{\mu}$ and all sufficiently large n

$$\begin{aligned} &\frac{c_n(\mu s)}{\mu} - \langle s, \gamma\rangle \\ &\quad \le \frac{1}{a_n\mu} \log\left\{\exp(a_n |\mu| \varepsilon) + \exp\left[-a_n\mu\langle s, \gamma\rangle + \frac{a_n(c_n(2\mu s) - M(\varepsilon))}{2}\right]\right\}. \end{aligned} \tag{37}$$

For $-\bar{\mu} \le \mu < 0$, the sense of the inequality is reversed.

Below we prove that for all $0 < \varepsilon \le \bar{\varepsilon}$ there exist $\bar{\mu} = \bar{\mu}(\varepsilon) \in J_s$ and $\bar{n} = \bar{n}(\varepsilon, \bar{\mu})$ such that for all $|\mu| \le \bar{\mu}$ and all $n \ge \bar{n}$

$$|\mu| \varepsilon \ge -\mu\langle s, \gamma\rangle + \frac{c_n(2\mu s) - M(\varepsilon)}{2}. \tag{38}$$

If we accept this, then for $0 < \varepsilon \le \bar{\varepsilon}$, $0 < \mu \le \bar{\mu}$ and $n \ge \bar{n}$ we have

$$\frac{c_n(\mu s)}{\mu} - \langle s, \gamma \rangle \le \varepsilon + \frac{\log 2}{a_n \mu} \tag{39}$$

while for $0 < \varepsilon \le \bar{\varepsilon}$, $-\bar{\mu} \le \mu < 0$, and $n \ge \bar{n}$

$$\frac{c_n(\mu s)}{\mu} - \langle s, \gamma \rangle \ge -\varepsilon + \frac{\log 2}{a_n \mu}. \tag{40}$$

Taking $n \to \infty$, then $\mu \to 0$, then $\varepsilon \to 0$, we conclude

$$\langle s, \gamma \rangle \le D_s^- c(0) \le D_s^+(0) \le \langle s, \gamma \rangle. \tag{41}$$

This is (33).

For $0 < \varepsilon \le \bar{\varepsilon}$ there exists $\bar{\mu} = \bar{\mu}(\varepsilon) \in J_s$, and for all $|\mu| \le \bar{\mu}$ we have

$$|\mu|\varepsilon \ge -\mu\langle s, \gamma \rangle + c(2\mu s)/2 - M(\varepsilon)/4. \tag{42}$$

This holds since c is continuous at 0 and $M(\varepsilon) > 0$. Now (38) follows since $\{c_n(2\mu s)\}$ converges uniformly to $c(2\mu s)$ on the interval $|\mu| \le \bar{\mu}$ [Rockafellar (1970; Theorem 10.8)]. □

Acknowledgements. I have had many useful discussions about large deviations with Larry Brown and S. R. S. Varadhan. Varadhan suggested Gärtner's paper and Lemma III.2.

REFERENCES

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

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

BERMAN, A. and PLEMMONS, R. J. (1979). *Nonnegative Matrices in the Mathematical Sciences.* Academic, New York.

DACUNHA-CASTELLE, D. (1979). Formule de Chernoff pour une suite de variables réelles. In: *Grandes Déviations et Applications Statistiques,* 19–24. *Astérisque* **68.** Soc. Math. de France.

DONSKER, M. D. and VARADHAN, S. R S. (1975). Asymptotic evaluation of certain Markov process expectations for large time, I. *Commun. Pure Appl. Math.* **28** 1–47.

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

EKELAND, I. and TEMAN, R. (1976). *Convex Analysis and Variational Problems.* North Holland, Amsterdam.

ELLIS, R. S. (1984). *Entropy, Probability Theory, and Statistical Mechanics.* In preparation. To be published by Addison-Wesley, Reading, Mass.

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

KATO, T. (1966). *Perturbation Theory for Linear Operators.* Springer Verlag, New York.

LANFORD, O. E. (1973). Entropy and equilibrium states in classical statistical mechanics. In: *Statistical Mechanics and Mathematical Problems,* 1–113. Lecture Notes in Physics, Number 20. Edited by A. Lenard. Springer Verlag, Berlin.

ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.
SPITZER, F. L. (1971). *Random Fields and Interacting Particle Systems*. Math. Assoc. of America, Washington.
STEINEBACH, J. (1978). Convergence rates of large deviation probabilities in the multidimensional case. *Ann. Probab.* **6** 751–759.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003