The Annals of Probability
1992, Vol. 20, No. 3, 1147–1166

LARGE DEVIATIONS FOR EXCHANGEABLE RANDOM VECTORS

BY I. H. DINWOODIE[1] AND S. L. ZABELL

Northwestern University

Say that a family $\{P_\theta^n: \theta \in \Theta\}$ of sequences of probability measures is *exponentially continuous* if whenever $\theta_n \to \theta$, the sequence $\{P_{\theta_n}^n\}$ satisfies a large deviation principle with rate function λ_θ. If Θ is compact and $\{P_\theta^n\}$ is exponentially continuous, then the mixture

$$P^n(A) =: \int_\Theta P_\theta^n(A)\, d\mu(\theta)$$

satisfies a large deviation principle with rate function $\lambda(x) =: \inf\{\lambda_\theta(x): \theta \in S(\mu)\}$, where $S(\mu)$ is the support of the mixing measure μ. If $X_1, X_2, \dots$ is a sequence of i.i.d. random vectors, $\{\overline{X}_n\}$ the corresponding sequence of sample means and $P_\theta^n =: P_\theta \circ \overline{X}_n^{-1}$, then $\{P_\theta^n\}$ is exponentially continuous if the classical rate function $\lambda_\theta(v)$ is jointly lower semicontinuous and a uniform integrability condition introduced by de Acosta is satisfied. These results are applied in Section 4 to derive a large deviation theory for exchangeable random variables; the resulting rate functions are typically nonconvex. If the parameter space Θ is not compact, then examples can be constructed where a full large deviation principle is not satisfied because the upper bound fails for a noncompact set.

1. Introduction. Let $\{P^n: n \geq 1\}$ be a sequence of probability measures on the Borel σ-algebra B of a topological space X and let $\lambda: X \to [0, \infty]$ be a nonnegative function on X. The sequence $\{P^n\}$ is said to satisfy a *large deviation principle* with *rate function* λ [see, e.g., Varadhan, (1984); Ellis, (1985); Deuschel and Stroock (1989)] if λ is a lower semicontinuous function having the property that

$$\liminf_{n \to \infty} \frac{1}{n} \log P^n(U) \geq -\inf\{\lambda(x): x \in U\} \tag{1.1}$$

for every open set $U \subset X$; and

$$\limsup_{n \to \infty} \frac{1}{n} \log P^n(C) \leq -\inf\{\lambda(x): x \in C\} \tag{1.2}$$

for every closed set $C \subset X$.

In this paper we will consider the case when the probability measures P^n may be represented as mixtures of simpler components. Thus, let Θ be a first countable topological space, let μ be a probability measure on Θ and for every $\theta \in \Theta$, let $\{P_\theta^n: n \geq 1\}$ be a sequence of probability measures on X such that

Received March 1990; revised November 1990.
[1]Supported under an NSF Graduate Fellowship.
AMS 1980 *subject classifications*. Primary 60F10; secondary 62F20.
Key words and phrases. Large deviations, exchangeable random variables, mixtures.

the map $\theta \to P_\theta^n(A)$ is a (Borel) measurable function on Θ for every integer n and measurable set A. Finally, given Θ, μ and $\{P_\theta^n\}$, let P^n denote the mixture

$$P^n(A) =: \int_\Theta P_\theta^n(A)\, d\mu(\theta). \tag{1.3}$$

We will be interested in describing the large deviation behavior of the sequence of mixtures $\{P^n\}$ in terms of the large deviation behavior of the component sequences $\{P_\theta^n\}$; and, in particular, to obtain an expression for a rate function for $\{P^n\}$ in terms of rate functions λ_θ for each $\{P_\theta^n\}$. For example, it is not hard to show and it follows as a special case of the theory developed here, that if the measure μ is concentrated on a finite set, then the rate function for the mixture reduces to the infimum of the rate functions for the component measures. That is, if $S(\mu)$ denotes the support of μ and

$$\lambda(x) =: \inf\{\lambda_\theta(x) : \theta \in S(\mu)\}, \tag{1.4}$$

then $\lambda(x)$ is a rate function for the mixture sequence $\{P^n\}$ when $S(\mu)$ is finite.

Example 1.1. Let $Y_1, Y_2, \dots$ be an i.i.d. sequence of Bernoulli trials with $P[Y_n = 1] = p$ and $P[Y_n = 0] = 1 - p$. In this case the rate function $\lambda_p(x)$ for the sequence $P[\bar{Y}_n \in A]$ is given by

$$\lambda_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}, \qquad 0 \le x \le 1, \tag{1.5}$$

and $\lambda_p(x) = \infty$ otherwise [see, e.g., Lanford (1973), pages 38–39; Azencott (1980), page 17]. If $X_1, X_2, \dots$ is an infinite exchangeable sequence of 0's and 1's and $S_n =: X_1 + X_2 + \cdots + X_n$, then by de Finetti's theorem [see, e.g., de Finetti (1937)], there exists a (unique) probability measure μ on the unit interval such that

$$P[S_n = k] = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\, d\mu(p); \tag{1.6}$$

that is, the general 0–1 valued exchangeable sequence may be represented as a mixture of Bernoulli processes. If, for example, $S(\mu) = \{1/3, 4/5\}$, then it follows from (1.4), (1.5) and (1.6) that the rate function λ for $\{S_n/n\}$, the sequence of sample proportions arising from the mixture, is the infimum of $\lambda_{1/3}$ and $\lambda_{4/5}$; see Figure 1 below. [Note that λ does not depend on the values of $\mu(1/3)$ and $\mu(4/5)$.] This example is discussed further in Example 4.1.

As soon as the support of the measure μ contains a point which is not isolated, it becomes necessary to impose some continuity condition on the sequences $\{P_\theta^n\}$. A natural condition to impose is a stability property which we will term *exponential continuity*:

(1.7) Whenever $\theta_n \to \theta$, the sequence $\{P_{\theta_n}^n\}$ satisfies a large deviation principle with rate function λ_θ.

If $\{P_{\theta_n}^n\}$ is exponentially continuous, then a complete theory for its mixtures may be derived when Θ is compact. Such a theory is developed in Section 2.

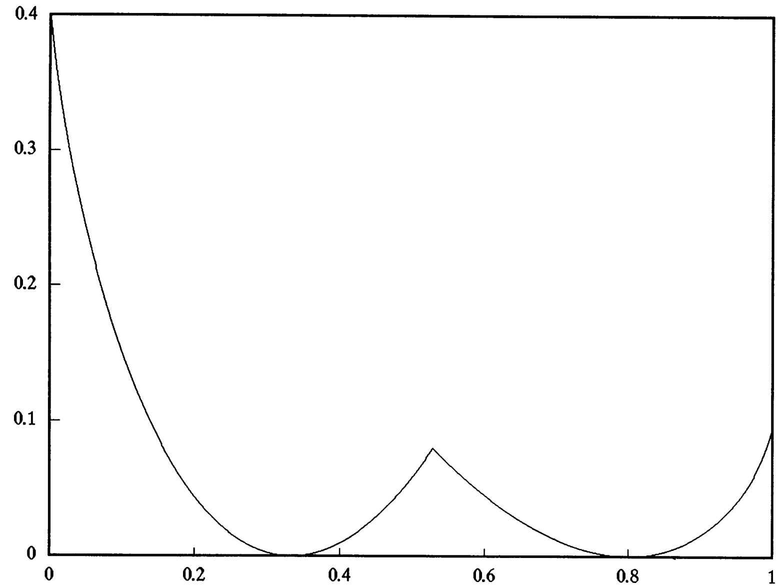

FIG. 1. *Graph of the rate function* λ *for the sequence of sample proportions arising from a mixture of Bernoulli trials with probabilities* $1/3$ *and* $4/5$.

In the case of large deviations of sums of i.i.d. random vectors, exponential continuity was shown to hold in certain cases by Bolthausen (1984). In Section 3, Bolthausen's theorem is extended to a much wider class (Theorem 3.1). An important distinction arises between the lower bound for open sets and the upper bound for compact and closed sets: while the lower bound holds in complete generality, the upper bound holds for compact sets if and only if $\lambda_\theta(x)$ is jointly lower semicontinuous and the upper bound for all closed sets holds if a uniform integrability assumption introduced by de Acosta (1985a) is satisfied. Example 3.1 illustrates that the upper bound can fail in the absence of this uniform integrability condition.

Some further results are given in Section 4. The previously published upper bound for sums of exchangeable random variables [de Acosta (1985a)] emerges as the lower convex hull of the true rate function (Proposition 4.1); that is to say, the best upper bound possible within the convex methodology that is the focus of de Acosta's paper. Example 4.2 illustrates that the upper bound need not hold if Θ is not compact. Finally, in the special case of a real-valued exchangeable sequence, Theorem 4.2 shows that the upper bound holds for all closed sets without any restriction on Θ or the underlying distribution.

2. Large deviation rates for general mixtures. Given a function $\lambda\colon \Theta \times X \to [0, \infty]$, let $\Lambda(\theta, A) = \inf\{\lambda(\theta, x)\colon x \in A\}$. We then have the following basic lower bound.

THEOREM 2.1. *Let $A \subset X$ be measurable. If*

$$\liminf \frac{1}{n} \log P_{\theta_n}^n(A) \geq -\Lambda(\theta, A) \tag{2.1}$$

whenever $\theta_n \to \theta \in \Theta$, then

$$\liminf \frac{1}{n} \log P^n(A) \geq -\inf\{\Lambda(\theta, A): \theta \in S(\mu)\}. \tag{2.2}$$

PROOF. Let $(\theta, v) \in \Theta \times A$. We must show that

$$\liminf \frac{1}{n} \log P^n(A) \geq -\lambda(\theta, v),$$

but may assume that $\lambda(\theta, v) < \infty$.

First, note that for every $\varepsilon > 0$, there exists an open set U_θ containing θ and an integer N_θ such that for every $\gamma \in U_\theta$,

$$P_\gamma^n(A) > \exp(-n[\Lambda(\theta, A) + \varepsilon]), \qquad n \geq N_\theta.$$

For if not, then in every neighborhood U_i from a countable neighborhood base at θ, we could find an element $\theta_i \in U_i$ and a number n_i arbitrarily large (in particular, such that $n_i > n_{i-1}$) with

$$P_{\theta_i}^{n_i}(A) \leq \exp(-n_i[\Lambda(\theta, A) + \varepsilon]).$$

But the sequence $\{\theta_i\}$ converges to θ, contradicting assumption (2.1).

Since U_θ is open, it has positive μ-measure and so for $n \geq N_\theta$,

$$\begin{aligned} P^n(A) &= \int_\Theta P_\theta^n(A)\, d\mu(\theta) \\ &\geq \int_{U_\theta} P_\theta^n(A)\, d\mu(\theta) \\ &\geq \int_{U_\theta} e^{-n[\Lambda(\theta, A) + \varepsilon]}\, d\mu(\theta), \end{aligned}$$

by the construction of U_θ. Hence

$$\liminf \frac{1}{n} \log P^n(A) \geq -\Lambda(\theta, A) - \varepsilon \geq \lambda(\theta, v) - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, the theorem follows. □

A corresponding upper bound may also be deduced, provided it is also assumed that the support of μ is compact.

THEOREM 2.2. *Let $A \subset X$ be measurable. If $S(\mu)$ is compact and*

$$\limsup \frac{1}{n} \log P_{\theta_n}^n(A) \leq -\Lambda(\theta, A), \tag{2.3}$$

whenever $\theta_n \to \theta \in S(\mu)$, *then*

$$\limsup \frac{1}{n} \log P^n(A) \le -\inf\{\Lambda(\theta, A): \theta \in S(\mu)\}. \tag{2.4}$$

PROOF. If $\Lambda(A) = 0$, the theorem is true, so we may assume otherwise. We will prove the case where $\Lambda(A) < \infty$, the proof when $\Lambda(A) = \infty$ being similar. We also assume without loss of generality that $S(\mu) = \Theta$.

Let $\varepsilon > 0$ and let θ be any element of Θ. Arguing as in the proof of Theorem 2.1, it follows from assumption (2.3) that there exist an open set U_θ containing θ and an integer N_θ such that for every $\gamma \in U_\theta$,

$$P_\gamma^n(A) \le \exp(-n[\Lambda(\theta, A) - \varepsilon]), \qquad n \ge N_\theta.$$

The family of open sets $(U_\theta)_{\theta \in \Theta}$ covers Θ and hence there is a finite subcover $(U_{\theta_i})_{1 \le i \le k}$. Thus

$$\begin{aligned} P^n(A) &= \int_\Theta P_\theta^n(A)\, d\mu(\theta) \\ &\le \sum_1^k \int_{U_{\theta_i}} P_\theta^n\, d\mu(\theta) \\ &\le \sum_1^k e^{-n[\Lambda(\theta_i, A) - \varepsilon]} \mu(U_{\theta_i}), \end{aligned}$$

hence

$$\begin{aligned} \limsup \frac{1}{n} \log P^n(A) &\le \max_{1 \le i \le k} \{-\Lambda(\theta_i, A) + \varepsilon\} \\ &\le \max_{1 \le i \le k} \{-\Lambda(A) + \varepsilon\} = -\Lambda(A) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, the theorem follows. □

If the level sets $L^\alpha =: \{x: \lambda(x) \le \alpha\}$ of a rate function λ are compact for every $\alpha \ge 0$, we will say that λ is *proper*. (This important property is so useful it is often included in the definition of a rate function.) The following lemma will be needed to establish the propriety of rate functions in Theorem 2.3. Recall that a topological space X is said to be *regular* (T_3) if it is Hausdorff, and every closed set $C \subset X$ and point $x \notin C$ can be separated by a pair of disjoint open sets; an extended real-valued function $f: X \to [0, \infty]$ on a topological space X is said to be lower semicontinuous at a point $x \in X$ if for each $c < f(x)$, there exists a neighborhood U_x of x such that $f(y) > c$ for every $y \in U_x$; and lower semicontinuous (on X) if it is lower semicontinuous at every point $x \in X$. From now on, we will sometimes write $\lambda(\theta, x) = \lambda_\theta(x)$.

LEMMA 2.1. *Let Θ be a compact first countable topological space and suppose that for every $\theta \in \Theta$, the sequence $\{P_\theta^n\}$ satisfies a large deviation principle with proper rate function λ_θ. If X is regular, if $\lambda_\theta(x)$ is jointly lower* *semicontinuous and if the upper bound* (2.3) *holds for every closed set, then the function* $\lambda(x) = \inf\{\lambda_\theta(x): \theta \in S(\mu)\}$ *has compact level sets.*

PROOF. Suppose there exists an α, $0 \le \alpha < \infty$, such that L^α is not compact. Choose and fix $\varepsilon > 0$. Because L^α is not compact, there exists a net $\{(\theta_i, v_i): i \in I\} \subset \Theta \times X$ such that $\{v_i\} \subset L^\alpha$, $\{v_i\}$ has no convergent subnet and $\lambda(\theta_i, v_i) \le \alpha + \varepsilon$ for all i. Because Θ is compact and first countable, it is sequentially compact, hence there exists a subsequence $\{(\theta_{i_k}, v_{i_k}): k \ge 1\}$ such that $\theta_{i_k} \to \theta \in \Theta$. Choose and fix $\beta > \alpha + 3\varepsilon$. Because L_θ^β, the level set for λ_θ, is compact, it contains only a finite number of elements in $\{v_{i_k}\}$, for otherwise $\{v_{i_k}\}$ would contain a convergent subnet. Thus for all k sufficiently large, say $k \ge k_0$, we have $\{v_{i_k}: k \ge k_0\} \subset (L_\theta^\beta)^c$.

Let $C_0 = \{v_{i_k}: k \ge k_0\}$. Because the sequence $\{v_{i_k}\}$ contains no convergent subnet, C_0 is closed. Because X is regular and L_θ^β is compact, there exist disjoint open sets U_1 and U_2 such that $L_\theta^\beta \subset U_1$ and $C_0 \subset U_2$. Let C denote the closure of U_2; because $U_1 \cap U_2 = \varnothing$, clearly $C = \overline{U_2} \subset U_1^c \subset (L_\theta^\beta)^c$, and it then follows from (2.3) that for every increasing sequence $\{n_k: k \ge k_0\}$,

$$\limsup_{k\to\infty} \frac{1}{n_k} \log P_{\theta_{i_k}}^{n_k}(C) \le -\Lambda(\theta, C) \le -\beta. \tag{2.5}$$

But on the other hand, because $\lambda(\theta, x)$ is a rate function and $U_2 \supseteq C_0$, we can find an increasing sequence $\{n_k: k \ge k_0\}$ such that for all $k \ge k_0$,

$$\frac{1}{n_k} \log P_{\theta_{i_k}}^{n_k}(U_2) \ge -\Lambda\big(\theta_{i_k}, U_2\big) - \varepsilon \ge -\alpha - 2\varepsilon > -\beta + \varepsilon. \tag{2.6}$$

But the two inequalities (2.5) and (2.6) cannot both hold because, by construction, $U_2 \subset C$. □

THEOREM 2.3. *Let* Θ *be a compact first countable topological space, let* X *be a regular topological space and suppose that for every* $\theta \in \Theta$, $\{P_\theta^n\}$ *is a sequence of probability measures on* X *satisfying a large deviation principle with rate function* λ_θ. *If* (2.1) *and* (2.3) *hold for every open set and every closed set, respectively, and if* $\lambda_\theta(x)$ *is jointly lower semicontinuous in* θ *and* x, *then every mixture* $\{P^n\}$ *of the form* (1.3), *with mixing measure* μ, *satisfies a large deviation principle with rate function* $\lambda(x) = \inf\{\lambda(\theta, x): \theta \in S(\mu)\}$. *If the rate functions* $\lambda_\theta(x)$ *are proper for every* $\theta \in S(\mu)$, *then* $\lambda(x)$ *is also proper.*

PROOF. It follows immediately from Theorems 2.1 and 2.2 that (2.2) and (2.4) hold for every open and closed set, respectively. Since Θ is compact and $\lambda(\theta, x)$ jointly lower semicontinuous, it follows that $\lambda(x)$ is likewise lower semicontinuous. If the functions $\lambda_\theta(x)$ are proper, the propriety of $\lambda(x)$ then follows from Lemma 2.1. □

The exponential continuity conditions necessary for the applicability of Theorems 2.1 and 2.2 are known at present to hold in two special cases.

(i) Let $X = E$ be a separable Banach space, let $X_1, X_2, X_3, \dots$ be a sequence of E-valued random vectors defined on a measurable space (Ω, A) and let $\overline{X}_n =: (X_1 + \cdots + X_n)/n$. For each $\theta \in \Theta$, let P_θ denote a probability measure on (Ω, A), let $P_\theta^n = P_\theta \circ \overline{X}_n^{-1}$ and let $\pi_\theta = P_\theta \circ X_1^{-1}$. If the sequence $\{X_j\}$ is independent and identically distributed with respect to P_θ for every $\theta \in \Theta$, if π_θ varies continuously in the weak topology and if

$$\sup_\theta \int e^{t\|X_1\|}\, dP_\theta < \infty, \qquad 0 \le t < \infty, \tag{2.7}$$

then it follows from Bolthausen (1984) that the family $\{P_\theta^n\colon \theta \in \Theta\}$ is exponentially continuous (with respect to the classical Cramér–Chernoff rate function). It then follows from Theorems 2.1 and 2.2 that if Θ is compact, then any sequence $\{P^n\}$ which is a μ-mixture of the sequences $\{P_\theta^n\}$ satisfies a large deviation principle with proper rate function given by (1.4). If P denotes the μ-mixture of $\{P_\theta\}$, then $\{X_j\}$ is exchangeable with respect to P (i.e., cylinder set probabilities are invariant under permutations of the time index) and $P[\overline{X}_n \in A] = P^n[A]$. Thus Theorems 2.1–2.3 give the large deviation behavior for the sequence of sample means of an exchangeable sequence taking values in E, whenever the $\{P_\theta\}$ appearing in the de Finetti representation of P satisfy Bolthausen's condition (2.7).

(ii) Let $X = S$, a complete separable metric space and let $Z_1, Z_2, \dots$ be an exchangeable sequence of random elements defined on a probability space (Ω, A, P) and taking values in S. Let $\overline{X}_n =: (\delta_{Z_1} + \cdots + \delta_{Z_n})/n$ denote the empirical probability measure, where δ_z denotes the Dirac measure concentrating mass at z. It then follows from the de Finetti representation theorem that P can be represented as a μ-mixture of probability measures $\{P_\theta\colon \theta \in \Theta\}$ defined on (Ω, A), where Θ is a subset of the probability measures on S which is closed in the weak topology, and for each $\theta \in \Theta$, the sequence $\{Z_j\}$ is independent and identically distributed with respect to P_θ. If $P^n = P \circ \overline{X}_n^{-1}$ and $P_\theta^n = P_\theta \circ \overline{X}_n^{-1}$, then P^n is the μ-mixture of $\{P_\theta^n\colon \theta \in \Theta\}$. It is well known that for each θ, $\{P_\theta^n\colon n \ge 1\}$ satisfies a large deviation principle with proper rate function

$$\lambda_\theta(\nu) = \int_S \log\left(\frac{d\nu}{d\pi_\theta}\right) d\nu \quad \text{if } \nu \ll \pi_\theta \quad \text{and} \quad \log\left(\frac{d\nu}{d\pi_\theta}\right) \in L^1, \tag{2.8}$$

and $\lambda_\theta(\nu) = \infty$, otherwise. Baxter and Jain [(1988), Theorem 5] state that the family $\{P_\theta^n\}$ is exponentially continuous with respect to the family $\{\lambda_\theta\}$ given in (2.8) if the probability measures $\pi_\theta =: P_\theta \circ Z_j^{-1}$ vary continuously in the weak topology. It then follows from Theorems 2.1–2.3 that if Θ is compact, then the sequence $\{P^n\}$ satisfies a large deviation principle with proper rate function given by $\lambda(\nu) = \inf_{\theta \in S(\mu)}\{\lambda_\theta(\nu)\}$, where $\lambda_\theta(\nu)$ is given by (2.8).

Unfortunately, in finite-dimensional spaces condition (2.7) is very restrictive; and there are a number of simple cases in which Bolthausen's theorem in its original form cannot be employed to deduce the large deviation behavior of an exchangeable sequence from Theorems 2.1–2.3. (For example, any mixture of exponential random variables; see Example 3.3 below.) For this reason, Bolthausen's theorem is extended in the next section to include such cases. This extension is cast in the setting of a locally convex space E to include as a special case the Sanov example just discussed, but the argument would not be essentially simpler were E assumed to be finite-dimensional.

3. Exponential continuity for sums of random vectors. Let E be a locally convex Hausdorff topological vector space, B the Borel σ-algebra of E and $X_1, X_2, X_3, \dots$ an infinite sequence of E-valued random vectors defined on a common measurable space (Ω, A). We assume there exists a complete, convex set $E_0 \subset E$, which is a Polish space in the relative topology and such that $X_n(\omega) \in E_0$ for every $n \geq 1$ and $\omega \in \Omega$. This assumption will insure that the random vectors $S_n =: X_1 + X_2 + \cdots + X_n$, and hence the sample means $\overline{X}_n =: S_n/n$ are A-measurable, and permits the application of certain results in Bahadur and Zabell (1979); this paper is cited below as BZ. It could in fact be weakened in one of several ways [see BZ, pages 591–592; Azencott (1980), pages 20–23; Csiszár (1984), pages 770–771], but it has the advantage of being relatively concrete and applicable to most examples of interest. (To be specific: Assumptions 1–3 of BZ would be sufficient for the remainder of this section except for Lemmas 3.1–3.2 and the necessity of lower semicontinuity in Theorem 3.1.)

Let Θ be a first countable topological space, for each $\theta \in \Theta$ let P_θ denote a probability measure on (Ω, A) and let $P_\theta^n = P_\theta \circ \overline{X}_n^{-1}$. Let E' denote the topological dual of E and for each $\theta \in \Theta$, define the function $\lambda_\theta\colon E \to [0, \infty]$ by

$$\lambda_\theta(v) = \sup_{\xi \in E'} \langle v, \xi \rangle - \log E_\theta[\exp\langle X_1, \xi \rangle].$$

Finally, let $\pi_\theta =: P_\theta \circ X_1^{-1}$; and let $\pi_{\theta_n} \Rightarrow \pi_\theta$ denote weak convergence of measures.

Theorem 3.1 employs two basic conditions, which will arise repeatedly in this section and the next:

$$\lambda\colon \Theta \times E \to [0, \infty] \text{ is lower semicontinuous;} \tag{3.1}$$

(3.2) $\{P_\theta^n\}$ is exponentially tight: For every $a > 0$, there exists a compact set $K_a \subset E$ such that $\sup_\theta\{P_\theta^n(K_a^c)\} < \exp(-na)$ for all n sufficiently large.

THEOREM 3.1. *Let $X_1, X_2, \dots$ be i.i.d. E-valued random vectors under P_θ. If the map $\theta \to P_\theta$ is continuous in the weak topology, then*:

(i) *For every open set $U \subset E$,*

$$\liminf \frac{1}{n} \log P_{\theta_n}^n(U) \geq -\Lambda(\theta, U). \tag{3.3}$$

(ii) *The upper bound*

$$\limsup \frac{1}{n} \log P^n_{\theta_n}(K) \leq -\Lambda(\theta, K) \tag{3.4}$$

holds for every compact set $K \subset E$ *if and only if* $\lambda(\theta, v)$ *is lower semicontinuous.*

(iii) *If* $\lambda(\theta, v)$ *is lower semicontinuous and* $\{P^n_\theta\}$ *is exponentially tight, then the upper bound* (3.4) *holds for every closed set* $C \subset E$.

PROOF. We first establish the lower bound (3.3). We may assume without loss of generality that $\Lambda(\theta, U) < \infty$ and that U is open and convex. It then follows that $\lim_{n\to\infty} n^{-1} \log P^n_\theta(U) = -\Lambda(\theta, U)$; see BZ, Theorem 2.3. Thus for n sufficiently large, say $n \geq N = N(\theta)$, we have $P^n_\theta(U) > 0$. Let $k \geq 1$ and $n \geq N + 1$. Because U is convex,

$$P^n_{\theta_n}(U) \geq \left(P^k_{\theta_n}(U)\right)^{[(n-N)/k]} P^{N+r_n}_{\theta_n}(U), \tag{3.5}$$

where the square brackets denote the greatest integer function and r_n is the remainder term defined by $n - N = k[(n-N)/k] + r_n$, $0 \leq r_n \leq k-1$.

Let ε be such that $0 < \varepsilon < \min_{0 \leq i \leq k-1} P^{N+i}_\theta(U)$. Because $P_{\theta_n} \Rightarrow P_\theta$ and U is open, there exists an integer $N_1 \geq 1$ such that if $n \geq N_1$, then

$$P^{N+r_n}_{\theta_n}(U) > \min_{0 \leq i \leq k-1} P^{N+i}_\theta(U) - \varepsilon = \delta > 0.$$

Thus by (3.5), we have

$$\begin{aligned}\liminf \frac{1}{n} \log P^n_{\theta_n}(U) &\geq \liminf \frac{1}{n} \log\left(P^k_{\theta_n}(U)\right)^{[(n-N)/k]}\\ &\geq \frac{1}{k} \log P^k_\theta(U).\end{aligned}$$

This is true for any k, hence taking the limit inferior of the right side we get

$$\liminf \frac{1}{n} \log P^n_{\theta_n}(U) \geq \liminf \frac{1}{k} \log P^k_\theta(U) = -\Lambda(\theta, U),$$

again applying Theorem 2.3 of BZ. (This elegant proof of the lower bound is due to Alejandro de Acosta, and is given here with his kind permission.)

We next turn to the upper bound (3.4). Let $\theta_n \to \theta$ and let $K \subset E$ be compact. We can assume that $\Lambda_\theta(K) > 0$. We will consider only the case where $\Lambda_\theta(K) < \infty$, the argument when $\Lambda_\theta(K) = \infty$ being quite similar. Let $\varepsilon > 0$, and for each $v \in K$, let B_v be an open convex, balanced neighborhood of $0 \in E$ such that $\Lambda_\theta(v + B_v) > \lambda_\theta(v) - \varepsilon$. Such a set exists by Theorem 3.2 of BZ. Let $N_v = v + \frac{1}{2}\overline{B}_v \subset v + B_v$, and let K_v be the closed, convex hull of $N_v \cap K$. Then K_v is compact because K is compact and K_v is closed and precompact in a complete subset of E. Since $K_v \subset N_v$, we see that

$$\Lambda_\theta(K_v) \geq \Lambda_\theta(N_v) \geq \Lambda_\theta(v + B_v) > \lambda_\theta(v) - \varepsilon \geq \Lambda_\theta(K) - \varepsilon.$$

Now K is compact and $K \subset \bigcup_{v \in K}(v + \frac{1}{2}B_v)$, so K has a finite subcover:

$$K \subset \bigcup_{i=1}^{k} \left(v_i + \tfrac{1}{2}B_{v_i}\right) \subset \bigcup_{i=1}^{k} K_{v_i}.$$

Let $K_i =: K_{v_i}$, $1 \le i \le k$. Now because each K_i is convex, for any $\gamma \in \Theta$,

$$\sup_{n \ge 1} \frac{1}{n} \log P_\gamma^n(K_i) = \limsup \frac{1}{n} \log P_\gamma^n(K_i) \le -\Lambda_\gamma(K_i);$$

the last inequality follows from Lemma 2.5 of BZ since K_i is compact. Thus

$$\frac{1}{n} \log P_{\theta_n}^n(K) \le -\Lambda_{\theta_n}\left(\bigcup_{i=1}^{k} K_i\right) + \log\left[\frac{k}{n}\right],$$

which gives immediately that

$$\limsup \frac{1}{n} \log P_{\theta_n}^n(K) \le -\liminf \Lambda_{\theta_n}\left(\bigcup_{i=1}^{k} K_i\right).$$

Since λ is assumed lower semicontinuous and $\bigcup_{i=1}^{k} K_i$ is compact, it follows that the map $\Lambda(\theta, \bigcup_{i=1}^{k} K_i)$ from Θ to $[0, \infty]$ is lower semicontinuous. Thus

$$\limsup \frac{1}{n} \log P_{\theta_n}^n(K) \le -\Lambda_\theta\left(\bigcup_{i=1}^{k} K_i\right) \le -\Lambda_\theta(K) + \varepsilon,$$

and since $\varepsilon > 0$ is arbitrary, the upper bound for compact sets follows.

Now suppose that (3.1) does not hold. We discuss only the case $\lambda_\theta(v) < \infty$; the case $\lambda_\theta(v) = \infty$ has a similar proof. Then there exists a pair $(\theta, v) \in \Theta \times E$ and an $\varepsilon > 0$ such that for every neighborhood $N_\theta \times U_v$ of (θ, v), there exists a point $(\gamma, w) \in N_\theta \times U_v$ with $\lambda_\gamma(w) < \lambda_\theta(v) - 2\varepsilon$. Since λ_θ is lower semicontinuous, there is a neighborhood U of v such that $\Lambda_\theta(U) > \lambda_\theta(v) - \varepsilon$. Let $\{N_k: k \ge 1\}$ be a countable open neighborhood base at θ and let $\{U_k\}$ be a countable open neighborhood base for E_0 at $v \in E$. We can assume that $U_k \subset U$. Let (γ_k, v_k) be an element of $N_k \times U_k$ such that $\lambda(\gamma_k, v_k) < \lambda(\theta, v) - 2\varepsilon < \Lambda_\theta(U) - \varepsilon$. Now since

$$\liminf \frac{1}{n} \log P_{\alpha_k}^n(U_k) \ge -\Lambda(\alpha_k, U_k) \ge -\lambda(\alpha_k, v_k) > -\Lambda_\theta(U) + \varepsilon,$$

there exists n_k such that if $n \ge n_k$, $(1/n)\log P_{\alpha_k}^n(U_k) > -\Lambda_\theta(U) + \varepsilon$. Since $P_{\alpha_k}^{n_k}$ is regular, there is a compact set $K_k \subset U_k$ such that

$$\frac{1}{n} \log P_{\alpha_k}^{n_k}(K_k) > -\Lambda_\theta(U) + \varepsilon.$$

Now let $K = \bigcup_k K_k \wedge \{v\}$, which is compact. Let $\theta_n = \alpha_k$ when $n \in [n_k, n_{k+1})$. Then along the subsequence $\{n_k\}$,

$$\frac{1}{n_k} \log P_{\theta_{n_k}}^{n_k}(K) \ge \frac{1}{n_k} \log P_{\alpha_k}^{n_k}(K_k) > -\Lambda_\theta(U) + \varepsilon,$$

and hence (3.4) does not hold for K.

Suppose finally that both (3.1) and (3.2) are satisfied. Then for any closed set C,

$$P^n_{\theta_n}(C) \le P^n_{\theta_n}(C \cap K_a) + P^n_{\theta_n}(K^c_a),$$

which implies that

$$\limsup \frac{1}{n} \log P^n_{\theta_n}(C) \le \max\{-\Lambda_\theta(C \cap K_a), -a\}$$
$$\le \max\{-\Lambda_\theta(C), -a\}.$$

Since $a > 0$ is arbitrary, the result follows. □

REMARK 3.1. In Theorem 3.1, the lower bound for open sets and the upper bound for compact sets do not require that $\phi_\theta(\xi) =: E_\theta[\exp\langle X_1, \xi\rangle]$ be everywhere finite in a neighborhood of the origin, and remain nontrivial provided $\phi_\theta(\xi) < \infty$ for some nonzero $\xi \in E'$; see for example, Lanford [(1973), pages 47–48] for a simple example of the phenomenon in the i.i.d. case.

The following two lemmas provide simple sufficient conditions for the lower semicontinuity and exponential tightness conditions (3.1) and (3.2). For the lower semicontinuity of $\lambda(\theta, v)$, the only explicit result of this nature that we have been able to find in the literature is that of Azencott and Ruget [(1977), page 12]; the first condition given in Lemma 3.1 is substantially less restrictive, and is close to best possible.

LEMMA 3.1. (i) *If for every sequence $\theta_n \to \theta \in \Theta$ and $\xi \in E'$, there exists a sequence (ξ_n) in E' such that $\xi_n \to \xi$ in the Mackey topology and*

$$\limsup_{n\to\infty} \log \phi_{\theta_n}(\xi_n) \le \log \phi_\theta(\xi), \tag{3.6}$$

then $\lambda(\theta, v)$ is lower semicontinuous.

(ii) *If E is reflexive and $\lambda(\theta, v)$ is lower semicontinuous, then* (3.6) *is satisfied.*

(iii) *If*

$$\sup_{\theta\in\Theta} \int_E e^{\xi(v)} \pi_\theta(dv) = M_\xi < \infty, \qquad \xi \in E', \tag{3.7}$$

then $h_\xi(\theta, v) := \langle v, \xi\rangle - \log \phi_\theta(\xi)$ is continuous for every $\xi \in E'$ and $\lambda(\theta, v)$ is jointly lower semicontinuous in θ and v.

PROOF. Let $(\theta, v) \in \Theta \times E$. We must show that for any $c < \lambda_\theta(v)$, there is a neighborhood $U_\theta \times U_v$ of (θ, v) such that for all $(\gamma, w) \in U_\theta \times U_v$, $\lambda_\gamma(w) > c$. We can assume that $v \in E_0$, since E_0 is closed and $\lambda_\theta(v) = \infty$ for all $v \notin E_0$. Furthermore, it is enough to show that for any sequence $(\theta_n, v_n) \to (\theta, v)$, $\liminf \lambda_{\theta_n}(v_n) \ge \lambda_\theta(v)$, since E_0 is a metric space.

Let $(\theta_n, v_n) \to (\theta, v)$, let $c < \lambda_\theta(v)$ and choose $\xi \in E'$ such that $\langle v, \xi\rangle - \log \phi_\theta(\xi) > c$. By hypothesis, there exists a sequence $\{\xi_n\} \subset E'$ such that

$\xi_n \to \xi$ in the Mackey topology and $\limsup \phi_{\theta_n}(\xi_n) \le \phi(\xi)$. Since $\xi_n \to \xi$ in the Mackey topology and E_0 is Polish, it follows that $\langle v_n, \xi_n\rangle \to \langle v, \xi\rangle$. But $\lambda_{\theta_n}(v_n) \ge \langle v_n, \xi_n\rangle - \log \phi_{\theta_n}(\xi_n)$, hence

$$\begin{aligned}\liminf \lambda_{\theta_n}(v_n) &\ge \langle v, \xi\rangle - \limsup \log \phi_{\theta_n}(\xi_n)\\ &\ge \langle v, \xi\rangle - \log \phi_\theta(\xi) > c,\end{aligned}$$

and it follows that $\lambda(\theta, v)$ is lower semicontinuous.

The second assertion of the lemma is an immediate consequence of Mosco [(1971), page 519, Theorem 1]. To establish the final assertion of the lemma, let $\theta_n \to \theta$. It suffices to show that for fixed $\xi \in E'$,

$$\phi_{\theta_n}(\xi) = \int_E e^{\xi(v)} \pi_{\theta_n}(dv) \to \int_E e^{\xi(v)} \pi_\theta(dv) = \phi_\theta(\xi).$$

Now let $g_k(v) = \min\{\exp^{\xi(v)}, e^k\}$, so that g_k is bounded and continuous. Then $|\phi_{\theta_n}(\xi) - \phi_\theta(\xi)|$ is bounded by

$$\int_{\{\xi(v)>k\}} e^{\xi(v)}\, d\pi_{\theta_n} + \left|\int_E g_k\, d\pi_{\theta_n} - \int_E g_k\, d\pi_\theta\right| + \int_{\{\xi(v)>k\}} e^{\xi(v)}\, d\pi_\theta \tag{3.8}$$

for every k. Now for any $\gamma \in \Theta$,

$$\int_{\{\xi(v)>k\}} e^{\xi(v)}\, d\pi_\gamma \le e^{-k} \int_{\{\xi(v)>k\}} e^{2\xi(v)}\, d\pi_\gamma \le e^{-k} M_{2\xi}.$$

Thus for any $\varepsilon > 0$, we can choose and fix k large enough that each end term in (3.8) is less than $\varepsilon/3$ and since π_{θ_n} converges weakly to π_θ, the middle term can also be made less than $\varepsilon/3$ for large n, depending possibly on k. The lower semicontinuity of λ in θ and v then follows from the continuity of $h_\xi(\theta, v)$. □

Given a subset $A \subset E$, the *Minkowski functional* of A is defined to be the function $q_A(v) =: \inf\{t > 0: v \in tA\}$, with $q_A(v) = \infty$ if $v \notin tA$ for any $t \ge 0$. If $tv \in A$ for $0 \le t \le 1$ whenever $v \in A$, then the subset is said to be *positively balanced*. Alejandro de Acosta (1985a) has introduced the following useful sufficient condition for exponential tightness:

There exists a convex, compact, positively balanced set $K \subset E$ such that

$$M =: \sup_{\theta\in\Theta} \int e^{q_K(X_1)}\, dP_\theta < \infty. \tag{3.9}$$

The next lemma shows that a generalization of the Bolthausen condition (2.7), employed by Baxter and Jain (1988) for another purpose, may be used to establish the de Acosta condition (3.9) and thus verify (3.2). The lemma is a generalization of Theorem 3.1 in de Acosta (1985a).

LEMMA 3.2. *Suppose that E_0 is positively balanced and that there exists a countable family of seminorms $\{p_i\colon i \ge 1\}$ generating the relative topology on*

E_0 at the origin. If $\{\pi_\theta: \theta \in \Theta\}$ is tight and for each $t \in \mathbf{R}$ and $i \geq 1$,

$$\sup_{\theta \in \Theta} \int_E e^{tp_i(X_1)}\, dP_\theta < \infty,$$

then there exists a convex, compact, positively balanced set K satisfying (3.9) and $\{P_\theta^n\}$ is exponentially tight.

PROOF. The proof uses the notation and ideas from Theorem 3.1 of de Acosta (1985a). Choose $\beta \in (0,1)$. Since $\{\pi_\theta: \theta \in \Theta\}$ is tight, there exists a compact set $K_m \subset E_0$ such that $\pi_\theta(K_m^c) < \beta^m$ for all $\theta \in \Theta$. Since E_0 is complete, we may assume that K_m is convex, positively balanced and $K_m \subset K_{m+1}$. We can assume that $p_{i+1} \geq p_i$. Let $\tau_{\theta,i}(t) = \pi_\theta\{v: p_i(v) > t\}$ and let $t_{m,\theta,i} = \inf\{t > 0: \tau_{\theta,i}(t) < \beta^m\}$. It can then be shown that $\lim_{m\to\infty}\{m^{-1}\sup_\theta(t_{m,\theta,i})\} = 0$ [see de Acosta (1985), page 555]; thus for every integer $i \geq 1$, there exists an integer $m_i \geq 1$ such that whenever $m \geq m_i$, $m^{-1}\sup_\theta t_{m,\theta,i} \leq i^{-1}$. We may assume that $m_{i+1} > m_i$. For any $m \geq 1$, let i be such that $m_i \leq m < m_{i+1}$ and define $B_{m,\theta} =: \{v \in E_0: p_i(v) \leq t_{m,\theta,i}\}$. Define the compact set K to be the closed, convex, positively balanced hull of the set $K^* =: \bigcup_{m,\theta} m^{-1}(K_m \cap B_{m,\theta})$. To show that K is compact, it is enough to show that K^* is totally bounded since E_0 is complete. If U is any neighborhood of $0 \in E$, then there is a neighborhood U_i of the form $U_i = \{v \in E: p_i(v) < (1/k)\}$ such that $U_i \cap E_0 \subset U \cap E_0$. Then $m^{-1}B_{m,\theta} \subset U_i \cap E_0$ for all $m \geq M =: \max(m_i, m_k)$. Thus $K^* \subset \bigcup_1^M K_m \cup U$ and since $\bigcup_1^M K_m$ is compact, K^* is totally bounded.

Next we prove that K satisfies (3.9); for this it suffices to show there exists a constant $c > 0$ such that

$$\pi_\theta\{v: q_K(v) > t\} \leq c\beta^t \quad \text{for all } t > 0 \text{ and all } \theta \in \Theta. \tag{3.10}$$

Given the definition of K, however, the verification of (3.10) is immediate; see de Acosta (1985a). Finally, to see that (3.9) implies the exponential tightness of $\{P_\theta^n\}$, let $a > 0$ and $K_a = (a + \log M)K$. Then

$$\begin{aligned} P_\theta^n(K_a^c) &= P_\theta\{q_K(S_n) > n(a \log M)\} \\ &\leq P_\theta\left\{\exp\left[\sum_{i=1}^n q_K(X_i)\right] > e^{-n(a+\log M)}\right\} \\ &= \left(E_\theta[\exp\{q_K(X_1)\}]\right)^n e^{-na}M^{-n} \leq e^{-na}, \end{aligned}$$

and the lemma follows. □

The following examples illustrate various aspects of Theorem 3.1 and Lemmas 3.1 and 3.2. The first example demonstrates that the lower semicontinuity condition (3.1) does not by itself ensure the upper bound (3.4) for all closed sets, even when E is finite-dimensional.

EXAMPLE 3.1. Let $\Theta = \{\theta: 0 \le \theta \le 1\}$, let $E = \mathbf{R}^2$, let π_0 denote the Cauchy distribution on $\mathbf{R}^1 \subset \mathbf{R}^2$, let $\pi_\theta = (1-\theta)\pi_0 + \theta\delta_{(\theta^{-1},\theta)}$, and let $X_1, X_2, \dots$ be a sequence of i.i.d. random vectors in $\mathbf{R}^2$ with common distribution π_θ. Consider the closed unbounded set

$$C = \left\{(x,y) \in \mathbf{R}^2: y \ge \frac{1}{2x^2}, x \ge 1\right\}.$$

Although λ satisfies (3.1), the upper bound (3.4) does not hold for the set C; that is, $-\Lambda_0(C) = \infty$, yet there exists a sequence $\theta_n \to 0$ for which

$$\limsup \frac{1}{n} \log P_{\theta_n}^n > -\Lambda_0(C) = \infty.$$

The lower semicontinuity of $\lambda(\theta, v)$ in θ for every v follows from Lemma 3.1. We show that (3.4) is violated for a sequence $\{\theta_n\} \subset \Theta_0$ with $\theta_n \to 0$. By Chernoff's theorem, $\Lambda_0(C) = \infty$. Let $\theta > 0$, let $U_\theta = \{(x,y): y > (1/2x^2), \theta x > 1\}$ and let

$$A_{n,\theta} =: \left\{X_j = (\theta^{-1}, \theta), 1 \le j \le [n\theta] + 1; \frac{n - [n\theta] - 1}{\theta} < S_n - S_{[n\theta]+1} \in \mathbf{R}^1\right\}.$$

It is easily verified that $A_{n,\theta} \subset \{\overline{X}_n \in U_\theta\} \subset \{\overline{X}_n \in C\}$, and it then follows from Chernoff's theorem and Stirling's formula that $\lim_{n\to\infty} n^{-1} \log P_\theta^n(C) = 0$. Thus for any $\varepsilon > 0$, there exists an integer n_θ such that for all $n \ge n_\theta$, $n^{-1} \log P_\theta^n(C) > -\varepsilon$. If $\theta_k \to 0$ ($\theta_k > 0$), one can choose an increasing sequence of integers $1 \le n_1 < n_2 < \cdots$ such that $n_k^{-1} \log P_{\theta_k}^{n_k}(C) > -2\varepsilon > -\infty = -\Lambda_0(C)$ for all $k \ge 1$; and thus (3.4) is not satisfied for every sequence $\theta_n \to 0$. □

The next example illustrates that the de Acosta condition (3.9) does not entail, and is thus independent of, the lower semicontinuity condition (3.1).

EXAMPLE 3.2. Let $E = \mathbf{R}$; let $\Theta = \{\theta: 0 \le \theta \le 1\}$; let $\pi_0 = \delta_0$; let π_1 be the exponential distribution on $\mathbf{R}$ with density e^{-x}, $x \ge 0$; and let $\pi_\theta = (1-\theta)\pi_0 + \theta\pi_1$, $0 \le \theta \le 1$. Although $\pi_\theta \to \pi_0$ as $\theta \to 0$, $\lambda_\theta(x)$ is not lower semicontinuous: Clearly, $\lambda_0(x) = \infty$ except when $x = 0$, while if $0 < \theta \le 1$, it is easily computed that for $x \ge 0$, $\lambda_\theta(x) = \sup_{-\infty<t<1}\{tx - \log[1 - \theta + \theta/(1-t)]\} \le \sup_{-\infty<t<1} tx \le x$, using Jensen's inequality to show that the logarithmic term is nonnegative. It is clear, however, that condition (3.9) is satisfied, since for all $t < 1$,

$$E_\theta[\exp(t\|x\|)] = (1-\theta) + \theta\frac{1}{1-t} \le 1 + \frac{1}{1-t}, \qquad 0 \le \theta \le 1.$$

The next example illustrates that although (3.7) is sufficient for (3.1) and (2.7) is sufficient for (3.9), in neither case is the first condition necessary for the second.

EXAMPLE 3.3. Let $\Theta = \{\theta: \theta_0 < \theta < \infty\}$, $E = \mathbf{R}^1$ and let π_θ be exponentially distributed, with density $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$. Then $\lambda_\theta(x) = \theta x - 1 - \log(\theta x)$, $x > 0$; and $\lambda_\theta(x) = \infty$, $x \le 0$. If $M_\theta(t)$ is the moment generating function, then $M_\theta(t) = \theta/(\theta - t)$, $t < \theta$, and $M_\theta(t) = \infty$, $t \ge \theta$. Thus $\lambda_\theta(x)$ is jointly lower semicontinuous in θ and x, but $M_\theta(t)$ clearly does not satisfy (2.7) or (3.7). Nevertheless, the de Acosta condition (3.9) is satisfied for any sequence $\theta_n \to \theta$: take $K = \{x: |x| \le \theta_0/2\}$, any $\theta_0 > 0$.

The next example illustrates that condition (3.7) in general entails neither (2.7) nor (3.9), even when Θ consists of a single point.

EXAMPLE 3.4 [Based on de Acosta (1985a), Example 3.2]. Let $E = l^2$, the Hilbert space of square-summable sequences $v = \langle x_1, x_2, \dots \rangle$; let $e_i = \langle \delta_{ij} \rangle$ denote the ith unit vector in l^2 (where δ_{ij} is the Dirac delta function); and let X be a random vector in E such that $P[X = a_n e_n] = p_n$, where $a_n = \log n$, $p_n = cn^{-2}$ and $c = 1/\zeta(2)$. If ξ is a linear functional on l^2, then by the Riesz representation theorem, there exists an element $\langle t_1, t_2, \dots \rangle \in l^2$ such that $\xi(v) = \sum_{n \ge 1} t_n x_n$. Thus $E[\exp(\xi(X))] = c\sum_{n \ge 1} n^{(t_n - 2)} < \infty$, since $t_n \to 0$, being square-summable. Now let q_K be the Minkowski functional of a compact set. Clearly, $\lim_{n \to \infty} q_K(e_n) = \infty$ (otherwise infinitely many e_n would lie in a compact set $\{v \in E: q_K(v) \le \alpha < \infty\}$, which is impossible because no subsequence of $\{e_n: n \ge 1\}$ is Cauchy). Thus $E[\exp(q_K(X))] = c\sum_{n \ge 1} n^{\{qK(e_n) - 2\}} = \infty$, so that (3.9) and hence (3.2) fails. Likewise $E[\exp t\|X\|] < \infty$ precisely when $t < 1$, so that (2.7) fails as well.

4. Large deviations for exchangeable sequences. Given Theorem 3.1, Theorems 2.1 and 2.2 of Section 2 may be immediately applied to derive lower and upper bounds for large deviations of the sample means of an infinitely exchangeable sequence. Because the random vectors X_i are assumed to take values in the Polish space $E_0 \subset E$, their distribution is known to satisfy the de Finetti representation theorem: the sequence may be represented as a mixture of independent and identically distributed random variables; see, for example, Aldous [(1985), pages 50–51].

THEOREM 4.1. *If $X_1, X_2, \dots$ is an infinitely exchangeable sequence of random vectors with mixing measure μ and $P^n(A) =: P(\overline{X}_n \in A)$, then for any open $U \subset E$,*

$$\liminf \frac{1}{n} \log P^n(U) \ge -\inf\{\lambda(\theta, v): \theta \in S(\mu), v \in U\}.$$

If Θ is a compact metric space and (3.1) *and* (3.2) *hold, then $\{P^n\}$ satisfies a large deviation principle with proper rate function $\lambda =: \inf\{\lambda_\theta: \theta \in S(\mu)\}$.*

PROOF. The lower bound follows from Theorems 2.1 and 3.1. If (3.1) holds, then by Lemma 3.1, $h_\xi(\theta, v)$ is jointly continuous in θ and v for fixed $\xi \in E'$;

thus $\lambda(\theta, v) = \sup\{h_\xi(\theta, v): \xi \in E'\}$ is jointly lower semicontinuous in θ and v. If (3.2) holds, then $\{P_\theta^n: n \geq 1\}$ satisfies a large deviation principle with rate function λ_θ [de Acosta, (1985b)]. The conclusion then follows from Theorems 3.1 and 2.3, because every topological vector space is regular. □

EXAMPLE 4.1. We return to Example 1.1 to illustrate the difference between the rate function λ in Theorem 4.1 and the upper rate function given in Example 1 of de Acosta (1985a). Let $S(\mu)$ be the two-point set $\{p_1, p_2\}$, $0 < p_1 < p_2 < 1$. Theorem 2.3 shows that $\lambda = \min(\lambda_{p_1}, \lambda_{p_2})$ is an upper and lower rate function. Thus $\lambda(x)$ is nonconvex, zero precisely at the two points $\{p_1, p_2\}$, and strictly positive everywhere else.

Let $\lambda_*(x)$ denote the rate function given in de Acosta [(1985a), Example 1]. A simple direct calculation yields that

$$\lambda_*(x) = x\log\frac{x}{p_1} + (1-x)\log\frac{1-x}{1-p_1}, \qquad 0 \leq x \leq p_1,$$

$$= x\log\frac{x}{p_2} + (1-x)\log\frac{1-x}{1-p_2}, \qquad p_2 \leq x \leq 1;$$

while $\lambda_*(x) = 0$ if $p_1 \leq x \leq p_2$ and $\lambda_*(x) = \infty$ if $x \notin [0,1]$. Thus λ_* is the lower convex envelope of the true rate function λ, agreeing with λ except on (p_1, p_2).

Example 4.1 provides a simple illustration of an important limitation inherent in the convex methodology: the upper bound it provides does not coincide with the actual rate when multiple phases exist. This phenomenon is in fact a general one: the upper rate function it specifies for exchangeable sequences is always the lower convex envelope of λ (that is, the greatest lower semicontinuous convex function less than or equal to λ), and thus the best possible convex upper bound.

PROPOSITION 4.1. *Let E be a separable Banach space. If Θ is second countable and*

$$\lambda_*(v) = (\log\phi)^*(v) = \sup_{\xi \in E'} \langle v, \xi\rangle - \log\phi(\xi),$$

where $\phi(\xi)$ is the essential supremum of $\phi_{(\cdot)}(\xi)$: $\Theta \to [0,\infty]$, then $\lambda_*(v)$ is the lower convex envelope of $\lambda(v)$.

PROOF. We claim first that $\phi = \sup_{\theta \in S(\mu)} \phi_\theta$ when Θ is second countable. For if $\{\theta_n\}$ is a sequence in $S(\mu)$ and $\theta_n \to \theta$, then $\liminf_{n\to\infty} \phi_{\theta_n}(\xi) \geq \phi_\theta(\xi)$. It follows that $\phi_\theta(\xi) \leq \phi(\xi)$, for otherwise $\mu\{\gamma: \phi_\gamma(\xi) > \phi_\theta(\xi) - \varepsilon\} = 0$ for $\varepsilon > 0$ sufficiently small. Thus $\sup_{\theta \in S(\mu)} \phi_\theta \leq \phi$.

But if Θ is second countable, then it also follows that $\phi \le \sup_{\theta \in S(\mu)} \phi_\theta$. For if $\phi(\xi) > M_\xi = \sup_{\theta \in S(\mu)} \phi_\theta(\xi)$ for some ξ, then $\mu\{\theta \in \Theta: \phi_\theta(\xi) > M_\xi\} > 0$. But this would imply that an element of this set lies in $S(\mu)$, which is impossible. Thus $\phi = \sup_{\theta \in S(\mu)} \phi_\theta$.

Next, we claim that $\lambda_* \le \lambda$. To see this, let $\theta \in S(\mu)$ and let φ^* denote the convex conjugate of the function φ. Then $\phi \ge \phi_\theta$, hence $\lambda_* = (\log \phi)^* \le (\log \phi_\theta)^* = \lambda_\theta$. Since $\theta \in S(\mu)$ was arbitrary, it follows that $\lambda_* \le \lambda$.

Finally, let φ be an arbitrary lower semicontinuous, convex function such that $0 \le \varphi \le \lambda$. Then for any $\theta \in S(\mu)$, $\varphi \le \lambda_\theta$, hence $\varphi^* \ge \lambda_\theta^* = \log \phi_\theta$. Thus $\varphi^* \ge \sup\{\log \phi_\theta: \theta \in S(\mu)\} = \log \phi$. Now since φ is lower semicontinuous and convex, $\varphi = \varphi^{**} \le (\log \phi)^* = \lambda_*$, hence λ_* is the greatest lower semicontinuous convex function below λ. □

If Θ is not compact, then in general the upper bound need not hold.

EXAMPLE 4.2. Let E be the topological vector space of finite signed measures on the interval $\mathbf{R}_+ = [0, \infty)$, endowed with the weak topology. Let $\Theta = \{1, 2, 3, \dots\}$ and define $\mu(k) = 6/\pi^2 k^2$, $k \ge 1$. Let Q_k denote the uniform probability measure on the Borel sets of the interval $I_k =: [2k - 1, 2k]$ and let $P_k = Q_k^\infty$ denote the corresponding product measure on $\mathbf{R}^\infty$. Let $T: \mathbf{R}_+ \to E$ be defined by $T(x) = \delta_x$, where δ_x is the point mass at x, let $\Omega = \mathbf{R}_+^\infty$ and let X_i: $\Omega \to E$ be defined by $X_i(\{x_1, \dots, x_i, \dots\}) = T(x_i)$. Consider the mixture $P^n(A) = \sum_{k=1}^\infty P_k(\overline{X}_n \in A)\mu(k)$. For each $k \in \Theta$, λ_k as computed with (2.8) is both an upper and lower rate function for the sequence of probabilities $P_k^n = P_k \circ \overline{X}_n^{-1}$ (BZ, Section 7), but we will show that $\lambda(v) =: \inf\{\lambda_k(v): k \in \Theta\}$ is not an upper rate function for the mixture $\{P^n\}$, by constructing a closed set $C \subset E$ with

$$\lim_{n \to \infty} \frac{1}{n} \log P^n(C) > -\inf_{v \in C} \lambda(v). \tag{4.1}$$

Let C_j, $j \ge 1$, be the set of discrete probability measures on $\mathbf{R}_+$ having no more than j mass points all of which lie in the interval $[2j - 1, 2j]$. Each set C_j is closed, hence $C = \bigcup_{j \ge 1} C_j$ is closed. Noting that $P_k(\overline{X}_n \in C_k) = 1$, $1 \le n \le k$, and that $P_k(\overline{X}_n \in C_k) = 0$, $n \ge k + 1$, it follows that

$$P^n(C) = \sum_{k=1}^\infty P_k(\overline{X}_n \in C_k)\mu(k) \ge \sum_{k=n}^\infty \mu(k)$$

$$= \frac{6}{\pi^2} \sum_{k=n}^\infty k^{-2} \ge \frac{6}{\pi^2} \int_n^\infty x^{-2}\, dx = \frac{6}{\pi^2} \frac{1}{n}.$$

Thus $\lim_{n \to \infty} n^{-1} \log P^n(C) = 0$. But it follows from (2.8) that for any $k \ge 1$

and any measure $\nu \in C$, $\lambda_k(\nu) = \infty$, since ν is discrete and Q_k is not. Thus $\Lambda(C) = \infty$ and (4.1) follows.

Note that in Example 4.2, the upper bound for closed sets does hold for the component sequences $\{P_k^n\}$ [see, e.g., Groeneboom, Oosterhoff and Ruymgaart (1979); BZ, Section 7].

The final result states that if attention is restricted to real-valued exchangeable random variables, then the upper bound is universally valid. Note that Θ is not assumed to be compact.

THEOREM 4.2. *If $E = \mathbf{R}$, then for every closed set C in $\mathbf{R}$,*

$$\limsup \frac{1}{n} \log P^n(C) \le -\inf\{\lambda_\theta(v) : \theta \in \Theta, v \in C\}.$$

Theorem 4.2 follows as an immediate consequence of the following lemma. Although very simple, we have been unable to find a reference for it in the literature [the closest is Azencott (1980), page 13, Theorem 2.3]. Note that the constant 2 is best possible (consider the example of the binomial distribution).

LEMMA 4.1. *Let $\{X_i\}$ be an i.i.d. sequence in $\mathbf{R}$ under the probability P_θ. If $C \subset \mathbf{R}$ is closed, then*

$$P_\theta(\bar{X}_n \in C) \le 2\exp(-n\Lambda_\theta(C)).$$

PROOF. Let $U = \{x \in \mathbf{R}: \lambda_\theta(x) > c\}$; U is open because λ_θ is lower semicontinuous and in fact U is either an open convex set or the union of two open convex sets. To see this, consider first the case where there exists some number x_0 such that $\lambda_\theta(x_0) = 0$. Then λ_θ is monotone increasing on $[x_0, \infty)$ and monotone decreasing on $(-\infty, x_0]$. Thus both $U_1 = U \cap (x_0, \infty)$ and $U_2 = U \cap (-\infty, x_0)$ are open and convex. Since $\Lambda_\theta(U) > 0$, $U = U_1 \cup U_2$. Next consider the case $\lambda_\theta > 0$. By Corollary 2.1 of BZ, either $\lambda_\theta(-\infty) = 0$ or $\lambda_\theta(\infty) = 0$. If $\lambda_\theta(-\infty) = 0$ and $x_1 < x_2$, then by the convexity of λ_θ,

$$\begin{aligned}\lambda_\theta(x_1) &= \lambda_\theta\left(\alpha x_2 + (1-\alpha)\left(\frac{x_1}{1-\alpha} - \frac{\alpha x_2}{1-\alpha}\right)\right) \\ &\le \alpha\lambda_\theta(x_2) + (1-\alpha)\lambda_\theta\left(\frac{x_1}{1-\alpha} - \frac{\alpha x_2}{1-\alpha}\right).\end{aligned}$$

Then let $\alpha \to 1$ to see that $\lambda_\theta(x_1) \le \lambda_\theta(x_2)$; thus λ_θ is monotone. A similar argument shows that λ_θ is monotone if $\lambda_\theta(\infty) = 0$. It follows that in either subcase λ_θ is monotone and hence U is convex.

Thus $U = U_1 \cup U_2$, where U_1 and U_2 are open convex, but one of U_1 or U_2 may be empty. Now by Lemma 2.2 and Theorem 2.1 of BZ,

$$\begin{aligned}\frac{1}{n}\log P_\theta(\bar{X}_n \in U) &\le \frac{1}{n}\log\big[P_\theta(\bar{X}_n \in U_1) + P_\theta(\bar{X}_n \in U_2)\big]\\ &\le \frac{1}{n}\log\big[e^{-n\Lambda_\theta(U_1)} + e^{-n\Lambda_\theta(U_2)}\big]\\ &\le \frac{1}{n}\log[2e^{-nc}] \le \frac{1}{n}\log 2 - c.\end{aligned}$$ □

REMARK 4.1. In addition to de Acosta (1985a), Bahadur and Raghavachari (1972), Bartfai (1979) and Izmirlian (1990) have studied large deviation questions involving exchangeable sequences. Bahadur and Raghavachari (1972) consider the problem of the asymptotic efficiency of tests, using a large deviation criterion for the optimality of a test and give as one illustration testing the composite null hypothesis of independence versus the composite alternative of exchangeability. Bartfai (1979) investigates a large deviation question of a very different type from that discussed here. If $X_1, X_2, \dots$ is a real-valued infinite exchangeable sequence, with $E|X_j| < \infty$ and $S_n =: X_1 + \cdots + X_n$, then it follows from the Birkhoff ergodic theorem that the random variable $Z =: \lim_{n\to\infty} S_n/n$ exists almost surely. Bartfai considers deviations of S_n/n from Z and states conditions under which exponential convergence of S_n/n to Z will occur. Izmirlian (1990) has studied large deviations for partially exchangeable sequences having a finite state space.

Acknowledgments. Our thanks to S. R. S. Varadhan, who suggested investigating the exponential continuity condition, to Alejandro de Acosta, for permission to quote here his generalization of Bolthausen's theorem in the case of the lower bound and to C. Ionescu-Tulcea for several helpful discussions.

REFERENCES

ALDOUS, D. J. (1985). Exchangeability and related topics. *Ecole d'Été de Probabilités de Saint-Flour XIII 1983. Lecture Notes in Math.* **1117**. Springer, New York.

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

AZENCOTT, R. and RUGET, G. (1977). Mélanges d'équations différentielles et grands écarts à loi des grands nombres. *Z. Wahrsch. Verw. Gebiete* **38** 1–54.

BAHADUR, R. R. and RAGHAVACHARI, M. (1972). Some asymptotic properties of likelihood ratios on general sample spaces. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 129–152. Univ. California Press, Berkeley.

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

BARTFAI, P. (1979). Remarks on the exchangeable random variables. *Publ. Math. Debrecen.* **26** 142–148.

BAXTER, J. R. and JAIN, N. C. (1988). A comparison principle for large deviations. *Proc. Amer. Math. Soc.* **103** 1235–1240.

BOLTHAUSEN, E. (1984). Large deviations in Banach spaces. *Ann. Probab.* **12** 427–435.

CSISZÁR, I. (1984). Sanov property, generalized I-projection, and a conditional limit theorem. *Ann. Probab.* **12** 768–793.

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

DE ACOSTA, A. (1985b). On large deviations of sums of independent random vectors. *Probability in Banach Spaces V. Lecture Notes in Math.* **1153** 1–14. Springer, New York.

DEUSCHEL, J.- D. and STROOCK, D. W. (1989). *Large Deviations.* Academic, San Diego.

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

DE FINETTI, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré Sect. B. (N.S.)* **7** 1–68.

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.* **7** 553–586.

IZMIRLIAN, G. (1990). Large deviations for additive functionals of a finite state partially exchangeable sequence. Unpublished manuscript.

LANFORD, O. E. (1973). Entropy and equilibrium states in classical statistical mechanics. *Statistical Mechanics and Mathematical Problems. Lecture Notes in Phys.* **20** 1–113. Springer, New York.

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

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

DEPARTMENT OF MATHEMATICS
LUNT HALL
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60208