PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 16, Fasc. 2 (1996), pp. 261–273

LARGE DEVIATIONS ON LINEAR SPACES*

BY

GEORGE L. O'BRIEN AND JIAMING SUN (NORTH YORK, ONTARIO)

Abstract. We discuss a method, which we call the *expansion method*, for proving large deviation principles and bounds. The method is applicable on general topological spaces but our main application is to prove a large deviation result for a sequence of random vectors taking values in a real locally convex linear space. As applications of this result, two general Cramér-type theorems are given. One comes directly from the main result; the proof of the other involves truncation and a continuity property of convex conjugation.

1. Introduction. In this paper, we prove some results about large deviations on real Hausdorff locally convex topological linear spaces. In the course of deriving these theorems, we employ two general methods for obtaining large deviation lower bounds.

Before describing our theorems we give some basic definitions which are applicable in the following more general context. Let E be a Hausdorff space and let $\mathscr{G}$, $\mathscr{F}$ and $\mathscr{K}$ denote the collections of open, closed and compact sets in E, respectively. Let (μ_n) be a sequence of probability (Borel) measures on E and let (α_n) be a sequence in $(0, 1]$ with $\alpha_n \to 0$.

1.1. DEFINITION. (a) We say (μ_n) satisfies a *narrow* (*vague*) *large deviation principle* (NLDP (VLDP)) with constants (α_n) and rate function $I\colon E \to [0, +\infty]$ if I is lower semicontinuous (lsc),

$$\limsup_{n\to\infty} \alpha_n \log \mu_n(F) \leqslant -\inf_{x\in F} I(x) \quad \text{for all } F \in \mathscr{F}, \tag{1.1}$$

$$\liminf_{n\to\infty} \alpha_n \log \mu_n(G) \geqslant -\inf_{x\in G} I(x) \quad \text{for all } G \in \mathscr{G} \tag{1.2}$$

(respectively, (1.2) holds as stated and (1.1) holds for all $F \in \mathscr{K}$).

* This research was supported by the Natural Sciences and Engineering Research Council of Canada.

(b) We refer (1.1) and (1.2) by themselves as a *large deviation upper bound* (LDUB) and a *large deviation lower bound* (LDLB), respectively.

For any probability measure μ and any $\alpha > 0$, we define the set function μ^α by $\mu^\alpha(A) = (\mu(A))^\alpha$.

1.2. Definition. Let μ_n and α_n be as above. We say the sequence $(\mu_n^{\alpha_n})$ is *equitight* if for all $\varepsilon > 0$ there is a $K \in \mathscr{K}$ such that $\mu_n^{\alpha_n}(K^c) < \varepsilon$ for all large n. Also, we say the rate function I: $E \to [0, +\infty]$ is *tight* if $\{x \in E: I(x) \leqslant y\}$ is compact for all $y \in (0, +\infty)$.

Equitightness is usually called *exponential tightness*, although it is generally expressed in logarithmic form. Our formulation emphasizes the similarity with the concept of tightness for probability measures. Equitightness is often an unattractive hypothesis, but it is not unduly restrictive in the sense that if E is Polish and (μ_n) satisfies an NLDP with tight rate function I, then $(\mu_n^{\alpha_n})$ must be equitight (cf. Lynch and Sethuraman [16] or O'Brien [17]).

Now suppose we have a collection Γ of continuous $[-\infty, +\infty)$-valued functions on E and suppose that for every $\xi \in \Gamma$ the *pressure*

$$\Psi(\xi) := \lim_{n\to\infty} \alpha_n \log \int_E \exp[\xi(x)/\alpha_n]\, \mu_n(dx) \tag{1.3}$$

exists in $(-\infty, +\infty]$. Then we call Ψ: $\Gamma \to (-\infty, +\infty]$ the *pressure function*. There are some situations in which the existence of the pressure function and the equitightness of $(\mu_n^{\alpha_n})$ together entail the existence of an NLDP with constants (α_n) and rate function I given by

$$I(x) := \Psi^*(x) := \sup\{\xi(x) - \Psi(\xi) \colon \xi \in \Gamma\}, \qquad x \in E. \tag{1.4}$$

Examples of this phenomenon can be found in Bahadur and Zabell [1] and Bryc [4].

An important theorem of large deviation theory is the *contraction principle* (Varadhan [21]) which can be stated loosely as "the continuous image of an NLDP is again an NLDP." More precisely, if E, (μ_n) and (α_n) are as above, if (1.1) and (1.2) hold for some tight rate function I and if λ: $E \to E_\lambda$ is a continuous map to some other Hausdorff space E_λ, then the sequence $(\mu_n \circ \lambda^{-1})$ of induced measures satisfies an NLDP with constants (α_n) and rate function I_λ given by

$$I_\lambda(y) = \inf\{I(x) \colon x \in E \text{ and } \lambda(x) = y\}, \qquad y \in E_\lambda. \tag{1.5}$$

Versions of this result with weaker continuity assumptions are given by Deuschel and Stroock [10] and O'Brien [17].

In Section 2, we develop an *expansion method*, which is in a sense an inverse of the contraction principle. Suppose we are given (μ_n) and (α_n) as above and a collection Λ of continuous functions λ: $E \to E_\lambda$, and suppose that we determine that, for each $\lambda \in \Lambda$, $(\mu_n \circ \lambda^{-1})$ satisfies an NLDP with constants (α_n) and rate function I_λ. If Λ is a rich enough class, we can deduce that (μ_n) also satisfies an NLDP with constants (α_n) and with rate function determined by the I_λ's. Note that it is important to allow the Hausdorff space E_λ to vary with λ.

The purpose of this method is much the same as that of the *projection method* used by Dawson and Gärtner [6] and Dembo and Zeitouni [9]. As a consequence some of the details of our proofs are related to some of theirs. The expansion method is worth presenting, however, since it is more general and more direct both in concept and in application.

Our application (in Section 3) of the expansion method is for the particular case when E is taken to be a real Hausdorff locally convex topological linear space V and Γ is taken to be the topological dual V^* of V. Gärtner [15] and Ellis [14] showed that if $V = R^d$ and if the pressure $\Psi(t)$ exists in $(-\infty, +\infty]$ for all $t \in V^*$, is finite in a neighbourhood of 0 and is "essentially smooth" (see our Section 3), then (μ_n) satisfies an NLDP with constants (α_n) and rate function given by (1.4). The finiteness requirement on Ψ automatically implies that $(\mu_n^{\alpha_n})$ is equitight in the case $V = R^d$. In this paper we present a similar theorem for general V as described above. Bryc [4] showed that for any metrizable V, if $(\mu_n^{\alpha_n})$ is equitight and Ψ is finite everywhere and sufficiently smooth, then the same conclusion holds. Various related theorems of the same type are given by Dembo and Zeitouni [9]. Their Corollary 4.5.27 (which is based on a result of Baldi [2]) and Corollary 4.6.14 require the pressure to be everywhere finite, while their Corollary 4.6.11 drops this requirement but at the cost of the extra assumption that V is the algebraic dual of another space. Also, we do not assume the pressure function to be lsc. Thus our result (Theorem 3.2) is different from theirs in several respects.

In Section 4, we apply our Theorem 3.2 and its proof to the case where μ_n is the distribution of the n'th sample mean from an independent and identically distributed (i.i.d.) sequence of V-valued random vectors. We get an NLDP if the equitightness condition holds. In the case where equitightness is not assumed, we develop a simple continuity property of convex conjugation and use it and a truncation argument to obtain a VLDP similar to Bahadur and Zabell's [1] extension of Cramér's theorem [5]. Thus, the Bahadur and Zabell theorem in its full generality can be deduced from the extended Gärtner–Ellis theorem. The continuity property is not directly related to Mosco convergence, as studied for example in Zabell [22]. In Section 5, we give a new sufficient condition for this VLDP to be in fact an NLDP. It is based on the contours of the rate function and is weaker than equitightness. It always holds for $V = R$.

2. The expansion method. The basic idea of this method was discussed in the Introduction. Here we first deal with the upper and lower bounds separately and then give an LDP result. This is followed by a discussion of the rate function. Let Λ be a collection of continuous maps $\lambda\colon E \to E_\lambda$. Let

$$\mathscr{H} := \{\lambda^{-1}(G_\lambda)\colon \lambda \in \Lambda,\ G_\lambda \text{ open in } E_\lambda\}.$$

We will call the topology generated by $\mathscr{H}$ the *minimal* topology. We obtain large deviation bounds for the original topology on E and for this possibly smaller topology. Define $\mu_{\lambda,n} := \mu_n \circ \lambda^{-1}$ for $n \in N$ and $\lambda \in \Lambda$.

2.1. PROPOSITION. *Suppose that, for each $\lambda \in \Lambda$, E_λ is a regular space and $(\mu_{\lambda,n})$ satisfies the LDUB with constants (α_n) and a rate function J_λ. If $K \subset E$ is compact under the minimal or original topology, then*

$$\limsup_{n\to\infty} \alpha_n \log \mu_n(K) \leqslant -\inf_{x\in K} J(x), \tag{2.1}$$

where J is the lsc function given by

$$J(x) = \sup_{\lambda\in\Lambda} J_\lambda(\lambda(x)). \tag{2.2}$$

If $(\mu_n^{\alpha_n})$ is equitight, then (2.1) *extends to all $K \in \mathscr{F}$.*

Proof. Since each J_λ is lsc, so is J. Let $\beta < \inf_{x\in K} J(x)$. Then for any $x \in K$ there exists a $\lambda \in \Lambda$ such that $J_\lambda(\lambda(x)) > \beta$, and so, by the lower semicontinuity and regularity, there exists an open set $B_\lambda \subset E_\lambda$ such that $\lambda(x) \in B_\lambda$ and

$$\limsup_{n\to\infty} \alpha_n \log \mu_{\lambda,n}(\bar{B}_\lambda) \leqslant -\inf_{y\in\bar{B}_\lambda} J_\lambda(y) < -\beta, \tag{2.3}$$

where $\bar{B}_\lambda$ is the closure of B_λ. So, by the compactness, there exist $x_k \in K$, $\lambda_k \in \Lambda$ and open $B_{\lambda_k} \subset E_{\lambda_k}$, $k = 1, \ldots, m$, such that $K \subset \bigcup_{k=1}^m \lambda_k^{-1}(B_{\lambda_k})$ and (2.3) holds for each k with λ and B_λ replaced by λ_k and B_{λ_k}. Therefore

$$\begin{aligned}\limsup_{n\to\infty} \alpha_n \log \mu_n(K) &\leqslant \bigvee_{k=1}^m \limsup_{n\to\infty} \alpha_n \log \mu_{\lambda_k,n}(B_{\lambda_k}) \\ &\leqslant \bigvee_{k=1}^m \big(-\inf_{y\in\bar{B}_{\lambda_k}} J_{\lambda_k}(y)\big) < -\beta,\end{aligned}$$

which yields (2.1). The extension to the closed sets is standard. ∎

2.2. PROPOSITION. *Suppose that every finite intersection of sets in $\mathscr{H}$ (defined above) is a union of sets in $\mathscr{H}$ (i.e., $\mathscr{H}$ is a base for the minimal topology, not just a subbase). Also assume that, for each $\lambda \in \Lambda$, $(\mu_{\lambda,n})$ satisfies an LDLB with constants (α_n) and rate function J_λ.*

(a) *If $G \subset E$ is open under the minimal topology, then*

$$\liminf_{n\to\infty} \alpha_n \log \mu_n(G) \geqslant -\inf_{x\in G} J(x), \tag{2.4}$$

where J is defined in (2.2).

(b) *Suppose in addition that $(\mu_n^{\alpha_n})$ is equitight and that for all distinct $x, y \in E$ there exist disjoint $H \in \mathscr{H}$ and $G \in \mathscr{G}$ such that $x \in H$ and $y \in G$. Then* (2.4) *holds for all $G \in \mathscr{G}$, and J is tight.*

Proof. (a) Let $G \subset E$ be open under the minimal topology and let $x \in G$. By the assumption about $\mathscr{H}$, there exist $\lambda \in \Lambda$ and open $G_\lambda \subset E_\lambda$ such that $x \in \lambda^{-1}(G_\lambda) \subset G$, so

$$\liminf_{n\to\infty} \alpha_n \log \mu_n(G) \geqslant \liminf_{n\to\infty} \alpha_n \log \mu_{\lambda,n}(G_\lambda)$$
$$\geqslant -\inf_{y\in G_\lambda} J_\lambda(y) \geqslant -J_\lambda(\lambda(x)) \geqslant -J(x).$$

(b) Let $G \in \mathscr{G}$ and $x \in G$ with $J(x) < +\infty$. Let $K \in \mathscr{K}$ be such that

$$\limsup_{n\to\infty} \mu_n^{\alpha_n}(K^c) \leqslant e^{-J(x)-1}. \tag{2.5}$$

For $y \in K\backslash G$, choose disjoint $H_y \in \mathscr{H}$ and $G_y \in \mathscr{G}$ with $x \in H_y$ and $y \in G_y$. Since finitely many such G_y's cover $K\backslash G$, the intersection H of the corresponding H_y's is disjoint from $K\backslash G$, so that $x \in H \subset G \cup K^c$. Therefore, by (a) we have

$$\liminf_{n\to\infty} \alpha_n \log \mu_n(H) \geqslant -J(x).$$

Combining this with (2.5) we see that also

$$\liminf_{n\to\infty} \alpha_n \log \mu_n(G) \geqslant -J(x).$$

The tightness of J is a consequence of the equitightness hypothesis about $(\mu_n^{\alpha_n})$ and (2.4). ■

Combining the above two propositions, we obtain:

2.3. THEOREM. *Assume that, for each $\lambda \in \Lambda$, E_λ is a regular space and that the set $\mathscr{H}$ satisfies the finite intersection condition described in Proposition* 2.2. *If, for each $\lambda \in \Lambda$, $(\mu_{\lambda,n})$ satisfies an NLDP with constants (α_n) and rate function J_λ, then*

(a) *(μ_n) satisfies the VLDP under the minimal topology with constants (α_n) and rate function J given by* (2.2), *and*

(b) *if, in addition, the hypotheses of Proposition* 2.2 (b) *hold, then (μ_n) satisfies the NLDP under the original topology with constants (α_n) and tight rate function J.*

We now specialize to the linear space situation described in the Introduction (so $E = V$). In this case we can say something about the form of the rate function J in Theorem 2.3. Define

$$J^*(\xi) = \sup\{\xi(x) - J(x):\ x \in V\}, \qquad \xi \in V^*.$$

2.4. COROLLARY. *Suppose that the conditions of Theorem* 2.3 (b) *hold. Then for any $\xi \in V^*$ the limit*

$$\underline{\Psi}(\xi) := \lim_{M\to\infty} \lim_{n\to\infty} \alpha_n \log \int_V \exp\{(\xi(x) \wedge M)/\alpha_n\} \mu_n(dx)$$

exists and $J^* = \underline{\Psi}$. Moreover, if J is convex, then $J = \underline{\Psi}^*$.

Proof. Since now (μ_n) satisfies an NLDP with constants (α_n) and rate function J, the result reduces to Theorem 3.1 of Dinwoodie [12]. ∎

2.5. Corollary. *Suppose that the conditions of Theorem* 2.3 (b) *hold and that for any* $\xi \in V^*$

$$\limsup_{n\to\infty} \alpha_n \log \int_V \exp(\xi(x)/\alpha_n)\,\mu_n(dx) < +\infty. \tag{2.6}$$

Then the pressure $\Psi(\xi)$ *exists for all* $\xi \in V^*$ *and* $J^* = \Psi$. *Moreover, if* J *is convex, then* $J = \Psi^*$.

Proof. It is easy to prove $\Psi = \underline{\Psi}$ in this case. ∎

Without (2.6) we cannot expect that $J^* = \Psi$ or $J = \Psi^*$. The following example shows us that even though the pressure function Ψ exists, $(\mu_n^{\alpha_n})$ is equitight and (μ_n) satisfies an NLDP with convex rate function J, J may not equal Ψ^*.

2.6. Example. Let $V = R^1$. Let $\mu_n\{k\} = k^{-2} e^{-nk}$ for all natural $k \geqslant n$ and let the remaining mass of μ_n be uniformly distributed on $[0, 1]$. Let $\alpha_n = 1/n$. Then obviously $(\mu_n^{1/n})$ is equitight and (μ_n) satisfies an NLDP with convex rate function given by

$$J(x) = \begin{cases} 0 & \text{if } 0 \leqslant x \leqslant 1, \\ +\infty & \text{otherwise.} \end{cases}$$

But $\Psi(t) = 0$ for $t < 0$, $\Psi(t) = t$ for $0 \leqslant t \leqslant 1$, and $\Psi(t) = +\infty$ otherwise. Therefore $\Psi^*(x) = J(x)$ for $x \leqslant 1$, but $\Psi^*(x) = x - 1 < J(x)$ for $x > 1$. However, $J = \underline{\Psi}^*$ by Corollary 2.4.

3. A general large deviation principle for linear spaces. We now apply the expansion method to obtain our extension of the Gärtner–Ellis theorem to infinite-dimensional linear spaces. Throughout this section, let V, V^*, (μ_n) and (α_n) be as in Section 1.

3.1. Definition. We say a convex function $f\colon V^* \to (-\infty, +\infty]$ is *essentially smooth* if for all $d \geqslant 1$ and $\xi_1, \ldots, \xi_d \in V^*$ the convex function $u(t_1, \ldots, t_d) := f\left(\sum_{i=1}^d t_i \xi_i\right)$ is essentially smooth on R^d (see the definition on p. 251 of Rockafellar [19]) and $0 \in \operatorname{int}\{t \in R^d\colon u(t) < \infty\}$.

When $V^* = R^d$, this definition reduces to the one of Rockafellar [19] except the former requires that $0 \in \operatorname{int}\{t \in R^d\colon f(t) < \infty\}$. We are now ready to state our theorem.

3.2. Theorem. *Assume that the pressure* $\Psi(\xi)$ *exists in* $(-\infty, +\infty]$ *for all* $\xi \in V^*$, *that* Ψ *is essentially smooth, and that* $(\mu_n^{\alpha_n})$ *is equitight. Then* (μ_n) *satisfies the NLDP with constants* (α_n) *and rate function* I *given by*

$$I(x) := \sup\{\xi(x) - \Psi(\xi);\ \xi \in V^*\}. \tag{3.1}$$

Note that $\Psi(0) = 0$ and Ψ is convex by Hölder's inequality, so the essential smoothness assumption is feasible. We apply the following known result (cf. Ellis [14] or De Acosta [7]) for our upper bound.

3.3. LEMMA. *Assume that $\Psi(\xi)$ exists in $(-\infty, +\infty]$ for all $\xi \in V^*$. Then for any $K \in \mathscr{K}$*

$$\limsup_{n\to\infty} \alpha_n \log \mu_n(K) \leqslant -\inf_{x\in K} I(x), \tag{3.2}$$

where $I(x)$ is defined in (3.1). *Furthermore, if $(\mu_n^{\alpha_n})$ is equitight, then* (1.1) *holds.*

Proof of Theorem 3.2. In view of Lemma 3.3, we need only to prove the lower bound half of Theorem 3.2. Take $E = V$ and the set Λ of Section 2 to be the set of all continuous linear maps from V into any Euclidean space; thus, for $\lambda \in \Lambda$, $E_\lambda = R^d$ for some $d \in N$ and λ has the form $\lambda(x) = (\xi_1(x), \ldots, \xi_d(x))$ for some $\xi_1, \ldots, \xi_d \in V^*$. For each such λ (and the corresponding d), we define the induced measures $\mu_{\lambda,n}$ as before. The corresponding pressure

$$\Psi_\lambda(t) := \lim_{n\to\infty} \alpha_n \log \int_{E_\lambda} \exp[\langle t, x\rangle/\alpha_n]\, \mu_{\lambda,n}(dx)$$

$$= \lim_{n\to\infty} \alpha_n \log \int_V \exp\Big[\Big(\sum_{k=1}^d t_k \xi_k(x)\Big)/\alpha_n\Big]\, \mu_n(dx) = \Psi\Big(\sum_{k=1}^d t_k \xi_k\Big)$$

exists for all $t = (t_1, \ldots, t_d) \in (R^d)^* = R^d$, where $\langle t, x\rangle$ is the Euclidean inner product. Since now $\Psi_\lambda(t)$ is essentially smooth with $0 \in \operatorname{int}\{t\colon \Psi_\lambda(t) < \infty\}$, $(\mu_{\lambda,n})$ satisfies an NLDP by the version of the Gärtner–Ellis theorem which does not require $\Psi_\lambda(t)$ to be lsc (for example, that of O'Brien and Vervaat [18]). In particular,

$$\liminf_{n\to\infty} \alpha_n \log \mu_{\lambda,n}(G_\lambda) \geqslant -\inf_{y\in G_\lambda} J_\lambda(y) \tag{3.3}$$

for every $\lambda \in \Lambda$ and G_λ open in R^d (with d determined by λ), where

$$J_\lambda(y) = \sup\{\langle y, t\rangle - \Psi_\lambda(t)\colon t \in R^d\}.$$

The class

$$\mathscr{H} := \{\lambda^{-1}(G_\lambda)\colon \lambda \in \Lambda,\ G_\lambda \text{ open in } R^d\}$$

is obviously closed under finite intersections and the topology generated by $\mathscr{H}$ is the weak topology on V and is, by the Hahn–Banach theorem, Hausdorff. We may therefore apply Proposition 2.2 (b) to conclude that

$$\liminf_{n\to\infty} \alpha_n \log \mu_n(G) \geqslant -\inf_{x\in G} J(x) \quad \text{for all } G \in \mathscr{G},$$

where

$$J(x) := \sup_{\lambda\in\Lambda} J_\lambda(\lambda(x)) = \sup_{\lambda\in\Lambda} \sup\{\langle \lambda(x), t\rangle - \Psi_\lambda(t)\colon t \in R^d\}$$

$$= \sup_{\lambda\in\Lambda} \sup\Big\{\sum_{k=1}^d t_k \xi_k(x) - \Psi\Big(\sum_{k=1}^d t_k \xi_k\Big)\colon t \in R^d\Big\} \leqslant I(x). \quad ■$$

3.4. Corollary. *Theorem* 3.2 *remains valid if the essential smoothness hypothesis is replaced by the assumption that the formula* (3.3) *holds for each* $\lambda \in \Lambda$.

The corollary follows from the above proof because the essential smoothnees was used only to prove (3.3). We will apply Corollary 3.4 in the proof of Theorem 4.4.

4. Large deviation principles for sample means. In this section, we take V and V^* as in Section 3. In order to avoid a complicated discussion of measurability (Bahadur and Zabell [1]), we *assume* further that μ_n is the distribution of the n'th sample mean

$$S_n := \frac{1}{n} \sum_{k=1}^{n} X_k$$

of some i.i.d. V-valued random sequence (X_n) defined on a probability space $(\Omega, \mathscr{B}, P)$ with common distribution μ. If μ is a probability measure concentrated on a Polish closed convex subset of V, then such a sequence (μ_n) can always be constructed with $\mu_1 = \mu$. We also take $\alpha_n := n^{-1}$ throughout the section. Then the pressure

$$\Psi(\xi) = \lim_{n\to\infty} \frac{1}{n} \log \int_V \exp[n\xi(x)]\, \mu_n(dx) = \log \int_V \exp[\xi(x)]\, \mu(dx)$$

exists in $(-\infty, +\infty]$ for all $\xi \in V^*$. We apply Theorem 3.2 to prove in Theorem 4.2 that (μ_n) satisfies a VLDP provided μ satisfies a certain regularity condition; and in Theorem 4.4 that (μ_n) satisfies an NLDP provided only that $(\mu_n^{1/n})$ is equitight. In the proof of Theorem 4.2, we use a truncation procedure. We remark that the same procedure was used by De Acosta et al. [8] in the case $V = R^d$, but with a fairly complex substitute for our Lemma 4.1.

Let us now equip V^* with the $\sigma(V^*, V)$ topology. We need not change the topology on V into the $\sigma(V, V^*)$ topology since lower semicontinuity of a convex function on V is equivalent for the two topologies (Ekeland and Temam [13]). Let f be a function from V (respectively, V^*) into $[-\infty, +\infty]$. The *convex conjugate* f^* of f is defined by

$$f^*(\xi) = \sup\{\xi(x) - f(x) : x \in V\}, \qquad \xi \in V^*$$

(respectively, $f^*(x) = \sup\{\xi(x) - f(\xi) : \xi \in V^*\}, x \in V$). In either case let $\operatorname{conv}(f)$ denote the greatest convex lsc function not exceeding f. Then (cf. Ekeland and Temam [13])

$$f \geqslant \operatorname{conv}(f) = f^{**} := (f^*)^*$$

and $f = f^{**}$ iff f is itself convex and lsc. If f is only assumed to be convex, then the greatest lsc function g not exceeding f can be shown to be convex also, and hence $g = \operatorname{conv}(f) = f^{**}$, so that, for any open set $G \subset V$ (V^*),

(4.1) $$\inf_{x\in G} f(x) = \inf_{x\in G} f^{**}(x).$$

We next give the advertised continuity property of convex conjugation.

4.1. LEMMA. *Let $\{f_\alpha;\ \alpha\in A\}$ be a set of functions from V (V^*) into $(-\infty, +\infty]$. Let*
$$g := \inf_{\alpha\in A} f_\alpha \quad \textit{and} \quad h := \sup_{\alpha\in A} f_\alpha.$$
Then:
(a) $g^* = \sup_{\alpha\in A} f_\alpha^*$.
(b) *If each f_α is convex and lsc, then $h^* = \phi^{**}$, where $\phi := \inf_{\alpha\in A} f_\alpha^*$.*
(c) *In case* (b), *if ϕ is also convex, then for all open $G \subset V^*$ (V)*

(4.2) $$\inf_{x\in G} \phi(x) = \inf_{x\in G} h^*(x).$$

Proof. (a) We only consider the case when each f_α is defined on V. The proof in the other case is similar. Since $g \leqslant f_\alpha$ for all α, $g^* \geqslant f_\alpha^*$ for all α, so $g^* \geqslant \sup_{\alpha\in A} f_\alpha^*$. If $g^*(\xi) > t \in R^1$, then $\xi(x) - g(x) > t$ for some $x \in V$. Therefore $\xi(x) - f_\alpha(x) > t$ for that x and some $\alpha \in A$, so that $f_\alpha^*(\xi) > t$ for that α. This proves (a).

(b) By (a) and the fact that $f_\alpha = f_\alpha^{**}$ for each α,
$$\phi^* = \sup_{\alpha\in A} f_\alpha^{**} = \sup_{\alpha\in A} f_\alpha = h.$$
Hence $h^* = \phi^{**}$.

(c) If ϕ is convex, (4.2) follows from (4.1). ■

4.2. THEOREM. *Let (μ_n) be defined as above. Suppose there exists an increasing sequence of compact convex sets $(K_l;\ l\in N)$ such that $\mu(K_l^c) \to 0$ as $l \to \infty$. Then (μ_n) satisfies the VLDP with rate function I given by*

(4.3) $$I(x) := \Psi^*(x) = \sup\{\xi(x) - \Psi(\xi):\ \xi \in V^*\}.$$

Proof. By Lemma 3.3, the inequality (1.1) holds for all compact sets. For $l \in N$, define
$$\mu_{l,n}(A) := P(S_n \in A;\ X_i \in K_l,\ i = 1, \ldots, n)\cdot\big(\mu(K_l)\big)^{-n}$$
for all Borel sets $A \subset V$. Then each $\mu_{l,n}$ is a probability measure with support in K_l. Define an lsc convex function Φ_l by
$$\Phi_l(\xi) = \log \int_{K_l} \exp[\xi(x)]\,\mu(dx) \quad \text{for all } \xi \in V^*.$$

Then the pressure

$$\Psi_l(\xi) := \lim_{n\to\infty} \frac{1}{n} \log \int_V \exp[n\xi(x)]\, \mu_{l,n}(dx) = \log \int_V \exp[\xi(x)]\, \mu_{l,1}(dx)$$

$$= \Phi_l(\xi) - \log \mu(K_l).$$

By the compactness of K_l, Ψ_l is everywhere finite and has everywhere finite directional derivatives. Thus equitightness and essential smoothness hold for $(\mu_{l,n}^{1/n})$ and Ψ_l. By Theorem 3.2, for any open set $G \subset V$ and for any $l \in N$,

$$\text{(4.4)} \quad \liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geqslant \liminf_{n\to\infty} \frac{1}{n} \log \mu_{l,n}(G) + \log \mu(K_l) \geqslant -\inf_{x\in G} \Phi_l^*(x).$$

By monotone convergence, $\Phi_l \uparrow \Psi$ and $\Phi_l^* \downarrow \phi := \inf_{l\in N} \Phi_l^*$ as $l \to \infty$. Since each Φ_l^* is convex, so is ϕ. By (4.4) and Lemma 4.1 (c), we conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geqslant -\inf_{x\in G} \phi(x) = -\inf_{x\in G} \Psi^*(x). \quad ■$$

4.3. Remark. The hypothesis about compact sets in Theorem 4.2 holds for all μ whenever V is Polish (i.e. separable and Fréchet), in particular when $V = R^d$. Briefly, the reason is that every such μ is tight and the closed convex hull of a compact set in such V is again compact. This argument is extended by Dembo and Zeitouni [9] to a more general situation in order to include the case where V is the space of finite signed measures on a Polish space E' and μ has support in the set $\mathcal{M}_1$ of probability measures on E'. The above convex hull property for $\mathcal{M}_1$ can be proved more easily as follows. By Prohorov's theorem (cf. Billingsley [3]), if $K \subset \mathcal{M}_1$ is compact, then K is tight, so its convex hull is tight, and hence is relatively compact. Therefore the VLDP holds in both cases by our Theorem 4.2.

4.4. THEOREM. *Let (μ_n) be defined as above. If $(\mu_n^{1/n})$ is equitight, then (μ_n) satisfies the NLDP with the rate function given in* (4.3).

Proof. By Corollary 3.4, we need only to verify (3.3). Let $\lambda = (\xi_1, \ldots, \xi_d)$ be given, where $\xi_1, \ldots, \xi_d \in V^*$. Let $Y_n := \lambda(X_n)$. Then (Y_n) is a sequence of i.i.d. R^d-valued random vectors, so, by Theorem 4.2 and Remark 4.3,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_{\lambda,n}(G_\lambda) = \liminf_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\sum_{k=1}^{n} Y_k \in G_\lambda\right) \geqslant -\inf_{x\in G_\lambda} I_\lambda(x)$$

for any open set $G_\lambda \subset R^d$, where $I_\lambda = \Psi_\lambda^*$ and

$$\Psi_\lambda(t) = \log \int_V \exp[\langle t, \lambda(x)\rangle]\, \mu(dx) = \Psi\left(\sum_{k=1}^{d} t_k \xi_k\right).$$

That is (3.3) holds. ■

4.5. Remark. Theorem 4.4 immediately implies a generalized Sanov's theorem under the weak topology (cf. Dembo and Zeitouni [9]). In fact, instead of assuming that the underlying space Σ is Polish in Sanov's theorem, we need only to assume that Σ is metrizable and the underlying measure is tight since the equitightness hypothesis can still be proved by Prohorov's theorem in this case.

5. A sufficient condition for an NLDP for sample means. Suppose throughout this section that we are in the same situation as at the beginning of Section 4 and that the condition in Theorem 4.2 holds, so that (μ_n) satisfies a VLDP with rate function given by (4.3). The objective of this section is to give a simple sufficient condition for the corresponding NLDP to hold. One sufficient condition is of course equitightness, which is useful in some important cases such as Sanov's theorem. A necessary and sufficient condition for the NLDP to hold is given in O'Brien [17], but it is not very easily verified. In the case $V = R^d$, when $d = 1$, the NLDP always occurs. Dinwoodie [11], building on an example of Slaby [20], has shown that this is not so for $d = 3$. The question is not yet settled for $d = 2$.

The condition given below is valid for all Polish spaces, but is mainly of interest for finite-dimensional spaces. It is based on the contours of I. For all $L > 0$, let $\Gamma_L := \{x \in V\colon I(x) \leqslant L\}$. Note the following facts. If $L_1 < L_2$, then $\Gamma_{L_1} \subset \Gamma_{L_2}$; each Γ_L is closed and convex since I is lsc and convex; every Γ_L is nonempty. Since every $\Gamma_L \neq \emptyset$ and I is convex, it follows that if Γ_L is compact for any $L > 0$, then all Γ_L's are compact. In this case the NLDP occurs. Our condition is a simple extension of this fact.

5.1. Definition. We will say I has *separated contours at infinity* if whenever $0 < L_1 < L_2$, there is a compact set $K \subset V$ and finitely many (possibly zero) pairs $(\xi_1, r_1), \ldots, (\xi_m, r_m)$ in $V^* \times R_1$ such that

$$\Gamma_{L_1} \subset \bigcap_{i=1}^{m} \{x \in E\colon \xi_i(x) \leqslant r_i\} \subset K \cup \Gamma_{L_2}. \tag{5.1}$$

5.2. Proposition. *Assume V is a Polish space. Let I have separated contours at infinity. Then (μ_n) satisfies the NLDP with the rate function I given in* (4.3).

Proof. By the VLDP we of course have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(A) \leqslant -\inf_{x\in A} I(x) \tag{5.2}$$

for compact A. Bahadur and Zabell [1] showed that the same result holds for all A expressible as a finite union of convex open sets. Our task is to prove (5.2) for all closed sets. So let F be a closed set with $\inf_{x\in F} I(x) > 0$. For any $0 < L_1 < L_2 < \inf_{x\in F} I(x)$, there are a compact set K and pairs $(\xi_1, r_1), \ldots, (\xi_m, r_m) \in V^* \times R^1$ such that (5.1) holds. Let

$$A := \bigcup_{i=1}^{m} \{x \in E\colon \xi_i(x) > r_i\}.$$

Since $F \subset \Gamma_{L_2}^c$, we have $F \subset (K \cap F) \cup A$, so that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\log \mu_n(F) &\leqslant \limsup_{n\to\infty} \frac{1}{n}\log \mu_n\big((K \cap F) \cup A\big)\\ &\leqslant \limsup_{n\to\infty} \frac{1}{n}\log \mu_n(K \cap F) \vee \limsup_{n\to\infty} \frac{1}{n}\log \mu_n(A)\\ &\leqslant \big(-\inf_{x\in K\cap F} I(x)\big) \vee \big(-\inf_{x\in A} I(x)\big) \leqslant \big(-\inf_{x\in F} I(x)\big) \vee (-L_1) = -L_1.\end{aligned}$$

Since $L_1 < \inf_{x\in F} I(x)$ is arbitrary, the proof is completed. ∎

If $E = R^1$, Γ_L must be a compact interval, a closed semi-infinite interval or R^1 itself; in each case all Γ_L must have the same form. Thus the hypothesis of Proposition 5.2 always holds. Cases where it holds for $E = R^2$ are when Γ_L is compact for some $L > 0$, when $\Gamma_L = F \times R^1$ for some closed convex $F \subset R^1$ for some $L > 0$, and when the boundaries of the Γ_L's are (branches of) hyperbolas with distinct asymptotes.

REFERENCES

[1] R. R. Bahadur and S. L. Zabell, *Large deviations of the sample mean in general vector spaces*, Ann. Probab. 7 (1979), pp. 587–621.

[2] P. Baldi, *Large deviations and stochastic homogenization*, Ann. Mat. Pura Appl. 151 (1988), pp. 161–177.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York 1968.

[4] W. Bryc, *Large deviations by the asymptotic value method*, in: *Diffusion Processes and Related Problems in Analysis*, Vol. 1, M. A. Pinsky (Ed.), Birkhäuser, Boston 1990, pp. 447–472.

[5] H. Cramér, *Sur un nouveau théorème-limite de la théorie des probabilités*, Actualités Sci. Indust. 736, *Colloque consacré à la théorie des probabilités*, 3, Hermann, Paris 1938, pp. 5–23.

[6] D. A. Dawson and J. Gärtner, *Large deviations from the Mckean-Vlasov limit for weakly interacting diffusions*, Stochastics 20 (1987), pp. 247–308.

[7] A. De Acosta, *Upper bounds for large deviations of dependent random vectors*, Z. Wahrsch. verw. Gebiete 69 (1985), pp. 551–565.

[8] — P. Ney and E. Nummelin, *Large deviation lower bounds for general sequences of random variables*, in: *Random walks, Brownian motion, and interacting particle systems*, R. Durrett and H. Kesten (Eds.), Birkhäuser, Boston 1991, pp. 215–221.

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, A. K. Peters, Wellesley, Mass. (formerly, Jones and Bartlett, Boston), 1993.

[10] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Academic Press, Boston 1989.

[11] I. H. Dinwoodie, *A note on the upper bound for i.i.d. large deviations*, Ann. Probab. 19 (1991), pp. 1732–1736.

[12] — *Identifying a large deviation rate function*, ibidem 21 (1993), pp. 216–231.

[13] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam 1976.

[14] R. S. Ellis, *Large deviations for a general class of random vectors*, Ann. Probab. 12 (1984), pp. 1–12.

[15] J. Gärtner, *On large deviations from the invariant measure*, Theory Probab. Appl. 22 (1987), pp. 24–39.

[16] J. Lynch and J. Sethuraman, *Large deviations for processes with independent increments*, Ann. Probab. 15 (1987), pp. 610–627.
[17] G. L. O'Brien, *Sequences of capacities, with connections to large-deviation theory*, J. Theoret. Probab. 9 (1996), pp. 19–35.
[18] – and W. Vervaat, *Compactness in large deviations theory*, Stochastic Process. Appl. 57 (1995), pp. 1–10.
[19] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton 1970.
[20] M. Slaby, *On the upper bound for large deviations of sums of iid random vectors*, Ann. Probab. 16 (1988), pp. 978–990.
[21] S. R. S. Varadhan, *Large Deviations and Applications*, SIAM, Philadelphia, 1984.
[22] S. L. Zabell, *Mosco convergence in locally convex spaces*, J. Funct. Anal. 110 (1992), pp. 226–246.

Department of Mathematics and Statistics
York University, 4700 Keele Street, North York, Ontario
Canada M3J 1P3
obrien@mathstat.yorku.ca & ysma9116@clid.yorku.ca

Received on 27.7.1995