The Annals of Probability
2000, Vol. 28, No. 3, 1259–1279

LARGE DEVIATION PROBABILITIES AND DOMINATING POINTS FOR OPEN CONVEX SETS: NONLOGARITHMIC BEHAVIOR

BY J. KUELBS

University of Wisconsin

The existence of a dominating point for an open convex set and a corresponding representation formula for large deviation probabilities are established in the infinite-dimensional setting under conditions which are both necessary and sufficient and follow from those used previously in $\mathbb{R}^d$. A precise nonlogarithmic estimate of large deviation probabilities applicable to Gaussian measures is also included.

1. Introduction. Let $X, X_1, X_2, X_3, \ldots$ be independent, identically distributed random vectors where $\mathscr{L}(X) = \mu$, and μ is a Borel probability measure on the real separable Banach space B. We assume throughout that the mean $m = \int_B x\, d\mu(x)$ exists as a Bochner integer. Let B^* denote the topological dual space of B,

$$\hat{\mu}(f) = \int_B e^{f(x)}\, d\mu(x), \qquad f \in B^* \tag{1.1}$$

and define

$$\lambda(x) = \sup_{f \in B^*}[f(x) - \log \hat{\mu}(f)], \qquad x \in B. \tag{1.2}$$

Then λ is a nonnegative convex rate function (possibly taking the value $+\infty$), and we define

$$\operatorname{dom}(\lambda) = \{x \in B\colon \lambda(x) < \infty\}. \tag{1.3}$$

If D is an open convex subset of B, then we will frequently assume that

$$m \notin \overline{D} \quad \text{and} \quad D \cap \operatorname{dom}(\lambda) \neq \phi, \tag{1.4}$$

and μ is such that

$$\int_B e^{t\|x\|}\, d\mu(x) < \infty \tag{1.5}$$

for all (or for some) $t > 0$.

If $S_n = X_1 + \cdots + X_n$ for $n \geq 1$, then if (1.5) holds for all $t > 0$, Donsker and Varadhan proved that the large deviation principle holds for $\{S_n/n\}$. More precisely, Theorem 5.3 of [7] implies that for any closed subset F of B,

$$\limsup \frac{1}{n} \log P(S_n \in nF) \leq -\inf_{x \in F} \lambda(x), \tag{1.6}$$

Received August 1998; revised November 1999.
AMS 1991 *subject classifications*. 60B12, 60F10.
Key words and phrases. Large deviation probabilities, dominating points for open convex sets, nonlogarithmic behavior.

and for any open subset G of B,

$$\liminf_n \frac{1}{n} \log P(S_n \in nG) \geq -\inf_{x \in G} \lambda(x). \tag{1.7}$$

If B is finite-dimensional, then the exponential integrability in (1.5) need only exist for some $t > 0$ instead of all $t > 0$.

In [1], limits of large deviation probabilities are obtained under far less restrictive conditions, but they hold only for convex sets (actually finite unions of convex sets). In particular, if D is an open convex subset of B, then Theorem 2.3 and Theorem 3.2 of [1] combine to imply that

$$\lim_n \frac{1}{n} \log P(S_n \in nD) = -\inf_{x \in D} \lambda(x). \tag{1.8}$$

In [1], no moments are assumed, and the argument proceeds via convexity considerations and subadditivity.

In [9] we proved that if D is an open convex subset of B such that (1.4) holds, and μ satisfies (1.5) for all $t > 0$, then (D, μ) has a unique *dominating point*. That is, there exists a point $a_0 \in \partial D$, the boundary of D, such that:

(1.9) (i) $\lambda(a_0) = \inf_{x \in D} \lambda(x) = \inf_{x \in \overline{D}} \lambda(x) < \infty$.
(1.9) (ii) For some $g \in B^*$ we have

$$D \subseteq \{x : g(x) \geq g(a_0)\}.$$

(1.9) (iii) $\lambda(a_0) = g(a_0) - \log \hat{\mu}(g)$, and
(1.9) (iv) $a_0 = \int_B x \exp\{g(x) - \log \hat{\mu}(g)\}\, d\mu(x)$, where the integral exists as a Bochner integral.

Of course, if $m \in \partial D$, then since $\lambda(m) = 0$ and $\lambda(\cdot) \geq 0$, we set $a_0 = m$, and (1.6) holds with g taken to be the zero linear functional. Hence the main interest is when $m \notin \overline{D}$, and assuming (1.5) for all $t > 0$, Theorem 1 in [9] proves dominating points exist in the infinite-dimensional setting with $g \in B^*$ satisfying a strict inequality in (1.9)(ii). A primary motivation for dominating points is the representation result

$$P(S_n \in nD) = \exp\{-n\lambda(a_0)\} J_n, \tag{1.10}$$

where

$$J_n = E\big(\exp\{-g(T_n)\} I_{\{T_n \in n(D - a_0)\}}\big), \tag{1.11}$$

$g(\cdot)$ is as in (1.9), $T_n = \sum_{j=1}^n (Z_j - a_0)$, and $Z, Z_1, Z_2, \ldots$ are i.i.d random vectors with $(d\mathscr{L}/d\mu)(Z)(x) = \exp\{g(x) - \log \hat{\mu}(g)\}$. The contrast between the Donsker–Varadhan results in (1.6) and (1.7) obtained under (1.5) for all $t > 0$, and those by Bahadur–Zabell in (1.8) without exponential moment assumptions, suggests that perhaps the results in [9] can be established under less restrictive conditions. Of course, when B is finite dimensional, the papers [10] and [11] are fundamental, and motivate the extensions we obtain here in the infinite-dimensional setting. This improvement is interesting, and perhaps somewhat surprising, since the assumption (1.5) for all $t > 0$ is near best possible for the full large deviation principle of (1.6) and (1.7) in the infinite-dimensional setting. Furthermore, our results are at the more delicate nonlogarithmic level, and the magnitude of J_n can estimated in some situations. For example, since $g(x) \geq 0$ for $x \in (D - a_0)$ by (1.9)(ii), it follows that J_n is obviously dominated by one, and (1.10) then immediately implies $P(S_n \in nD) \leq \exp\{-n\lambda(a_0)\}$ for all $n \geq 1$. An inequality of this general type also follows from the subadditive approach in [1], which gives $P(S_n \in nD) \leq \exp\{-n(\inf_{x\in D} \lambda(x))\}$, provided $n \geq k$ and k is such that $P(S_n \in nD) > 0$ for all $n \geq k$. With a bit more care, the subadditive approach actually yields this inequality for all $n \geq 1$. This can be seen, for example, by Lemma 2.2 of [3], or Lemma 3.1.2 and Lemma 3.1.3 in [6]. However, it does not identify $\inf_{x\in D} \lambda(x)$ as $\lambda(a_0)$ for some point $a_0 \in \overline{D}$. This identification is made in Lemma 2.1 below, provided the level sets of λ are weakly compact. Of course, the fact that there exists a dominating point a_0 such that the representation in (1.10) and (1.11) holds is a more delicate matter, and has proved useful in recent applications to the Gibbs conditioning principle in [4]. Further results and comments regarding the magnitude of J_n appear below.

If $A \subseteq B$, then $\mathrm{co}(A)$ denotes the convex hull of A. Throughout the paper, the set S denotes the topological support of μ. The closure of $\mathrm{co}(S)$ is given by $\overline{\mathrm{co}(S)}$ and Theorems 2.4 and 3.2 of [1] yield the fact that $\mathrm{dom}(\lambda) \subseteq \overline{\mathrm{co}(S)}$. Some important assumptions frequently invoked are as follows:

(A1) the level sets $\{x\colon \lambda(x) \leq t\}$ are weakly sequentially compact for all $t > 0$.
(A2) $\int_B e^{t\|x\|}\, d\mu(x) < \infty$ for some $t > 0$.

Since $\lambda(x)$ is the supremum of weakly continuous functions on B, it is lower semicontinuous in the weak topology on B (and hence also the norm topology). Thus the level sets in (A1) are weakly closed for all $t > 0$. In particular, if B is reflexive, then these level sets weakly compact provided they are bounded. Furthermore, if (A2) holds for some $t > 0$, then it is easy to see that the level sets are are all bounded in B. We also know from the Eberlein–Smulian theorem ([8], page 430) that these level sets are weakly sequentially compact iff they are weakly compact, but we prefer (A1) as stated.

If D is an open convex subset of B such that $a_0 \in \partial D$ and (1.9)(i) holds, then we call a_0 a *predominating point* for (D, μ). If the rate function λ has weakly sequentially compact level sets, then we will see every convex open subset of D satisfying $D \cap \mathrm{dom}(\lambda) \neq \emptyset$, has a predominating point.

To formulate our results it is useful to have easy access to another assumption, and hence we require some additional notation. If $f \in B^*$, we write μ^f to denote the Borel probability measure μ induces on $(-\infty, \infty)$ via the formula $\mu^f(A) = \mu(f^{-1}(A))$. Then $\widehat{\mu^f}(t) = \hat{\mu}(tf)$ for $t \in \mathbb{R}$ and $f \in B^*$, and we define

$$\mathrm{dom}(\widehat{\mu^f}) = \{t\colon \widehat{\mu^f}(t) < \infty\}. \tag{1.12}$$

The rate function for μ^f is defined by

$$\lambda_{\mu^f}(s) = \sup_{t\in\mathbb{R}}[ts - \log \widehat{\mu^f}(t)], \tag{1.13}$$

and if $\mathrm{dom}(\lambda_{\mu^f})$ has nonempty interior we write

$$\mathrm{int}(\mathrm{dom}(\lambda_{\mu^f})) = (a_f, b_f). \tag{1.14}$$

If μ^f has nondegenerate support, then the closed convex hull of its support, written $\overline{\mathrm{co}(\mathrm{supp}\ \mu^f)}$ has nonempty interior, and by Theorem 2.4c and Theorem 3.2 of [1], its closure is $[a_f, b_f]$. Furthermore, it is then well known that $\log \widehat{\mu^f}$ is strictly convex on $\mathrm{dom}(\widehat{\mu^f})$ (see the Remark below), and $\mathrm{dom}(\widehat{\mu^f})$ is an interval of $\mathbb{R}$, possibly degenerate, containing the origin. The left endpoint of $\mathrm{dom}(\widehat{\mu^f})$ is denoted by

$$t_f^- = \inf\{t: \widehat{\mu^f}(t) < \infty\}, \tag{1.15}$$

and the right endpoint by

$$t_f^+ = \sup\{t: \widehat{\mu^f}(t) < \infty\}. \tag{1.16}$$

Of course, when (A2) holds, we have for all $f \in B^*$ that

$$\mathrm{dom}(\widehat{\mu^f}) \supseteq (t_f^-, t_f^+) \text{ with } t_f^- < 0 \quad \text{and} \quad t_f^+ > 0. \tag{1.17}$$

The set $\mathrm{dom}(\widehat{\mu^f})$ may contain any combination of the endpoints t_f^- and t_f^+, and we are not assuming $\mathrm{dom}(\widehat{\mu^f})$ is open. Furthermore, if μ^f is assumed not to be concentrated at a single point, and either $t_f^- < 0$ or $t_f^+ > 0$, then an easy application of the Cauchy–Schwarz inequality shows the derivative of $\log \widehat{\mu^f}$ to be strictly increasing on (t_f^-, t_f^+). Hence $\log \widehat{\mu^f}$ is strictly convex there, and the limit in our assumption (A3) below exists. Our third assumption consists of the inequality

$$\lim_{t \uparrow t_f^+} \frac{d}{dt} \log \widehat{\mu^f}(t) \geq b_f. \tag{A3}$$

The following theorem presents our results regarding dominating points.

THEOREM 1. *Let D be an open convex subset of B such that* (1.4) *holds. Then we have the following results:*

(I) *If* (A1) *holds and $\lambda(x) = 0$ iff $x = m$, then (D, μ) has a predominating point a_0, that is, $a_0 \in \partial D$ and* (1.9)(i) *holds.*

(II) *Assume* (A1) *and* (A2) *hold. If D is a half-space, say $D = \{x: f(x) > c\}$, and $\widehat{\mu^f}$ satisfies* (A3)*, then (D, μ) has a dominating point d_0 given by the Bochner integral*

$$d_0 = \int_B x \exp(t_0 f(x))\, d\mu(x)/\hat{\mu}(t_0 f), \tag{1.18}$$

where $t_0 \in (0, t_f^+)$ and $g = t_0 f$ in (1.9).

(III) *Assume* (A1) *and* (A2)*. Then every open convex subset D of B satisfying* (1.4) *has a unique dominating point if and only if* (A3) *holds for all $f \in B^*$.*

REMARKS. 1. Once the open convex set is chosen, the proof of part (III) in Theorem 1 shows the assumption (A3) need only hold for some $f \in B^*$ satisfying (2.4) below. Also observe that if f satisfies (2.4), then (1.4) implies $\mathrm{co}(\mathrm{supp}(\mu^f))$ has nonempty interior.

2. If $\mathrm{co}(\mathrm{supp}(\mu^f))$ has empty interior, then μ^f is concentrated at a single point and (A3) holds trivially. Hence the assumption in (A3) is important only when the support of μ^f is nondegenerate.

3. If $t_f^+ = \infty$, then direct calculations imply (A3) if b_f is finite or infinite. When $t_f^+ < \infty$, then b_f is necessarily infinite, and hence (A3) would hold if $\log(\widehat{\mu^f})$ is essentially smooth. Conversely, if (A3) holds for both f and $-f$, then $\log(\widehat{\mu^f})$ is essentially smooth. Recall from [12] that if $h\colon \mathbb{R}^d \to \mathbb{R}$ and $\mathrm{dom}(h) = \{x \in \mathbb{R}^d\colon h(x) < \infty\}$, then h is said to be *essentially smooth* if:

(i) $\mathrm{int}(\mathrm{dom}(h)) \neq \emptyset$.
(ii) h is differentiable on $\mathrm{int}(\mathrm{dom}(h))$,
(iii) $\|Dh(x_i)\| \to \infty$ whenever $x_i \to x_0$, $x_0 \in \partial(\mathrm{dom}(h))$, and $\{x_i\} \subseteq \mathrm{int}(\mathrm{dom}(h))$.

Here $Dh(x)$ is the Frechet derivative of h at x. Also note that (iii) holds vacuously if $\mathrm{dom}(h) = \mathbb{R}^d$. Hence if $\hat{\mu}(f)$ is finite for all $f \in B^*$, then (A3) is satisfied. Essential smoothness can be defined analogously on B^*, and it is easy to show that this global assumption implies essential smoothness. Therefore (A3), for every $f \in B^*$. In particular, in $\mathbb{R}^d$ the assumption that (A3) holds for all $f \in B^*$ follows from the related assumptions used in [10] and [11].

The next theorem presents the representation result in (1.10). The proof is immediate.

THEOREM 2. *Let D be an open convex subset of B such that* (1.4) *holds. If* (D, μ) *has a dominating point, then* (1.10) *holds, where J_n is an in* (1.11).

Theorems 2 and 3 in [9] yield specific applications of the representation formula in (1.10) when (A2) holds for all $t > 0$. As mentioned previously, J_n in (1.11) is obviously dominated by one, but even more is true. Theorem 2 in [9] implies we have $J_n \le C_1 n^{-1/2}$ for some finite constant C_1, and when B is a Hilbert space, D is an open ball, and (1.4) holds with $m \notin \overline{D}$, then Theorem 3 in [9] implies there exists $c \in (0, 1)$ such that

$$c < n^{1/2} J_n < 1/c. \tag{1.19}$$

The upper bounds are the easy part of this work, whereas lower bounds are more delicate. Now we turn to the analogues of these results provided (A2) holds only for some $t > 0$.

In some special cases we are able to show $\lim_n n^{1/2} J_n$ exists and to identify the limit. We include this result when μ is Gaussian, but it extends to some special non-Gaussian cases as well. A similar calculation appeared recently in [4]. We also can obtain a precise value for the constant C_1 in the upper bound of J_n, and this is indicated in the remark following Proposition 2.

PROPOSITION 1. *Let μ be a mean zero Gaussian measure on a real separable Hilbert space B. Let D be the open ball $\{x: \|x-a\| < R\}$ where $0 < R < \|a\|$, and assume $a_0 \in \partial D$ is the dominating point for (D, μ). Let $x_0 = a - a_0$ and $b = 1/g(x_0)$, where g is as in* (1.9)(ii) *with strict inequality holding (this is always the case since the mean is zero and not in $\overline{D}$). Furthermore, let $G_2 = (Z-a_0) - g(Z-a_0)S(g)/\sigma_g^2$, where Z is as given following* (1.11), *$S(g) = E[(Z-a_0)g(Z-a_0)]$ and $\sigma_g^2 = g(S(g))$. Then G_2 is a centered Gaussian vector and*

$$\lim_n (2\pi\sigma_g^2 n)^{1/2} \exp\{n\lambda(a_0)\}\mu(\sqrt{n}D) = \int_0^\infty e^{-s} P(\|G_2\|^2 \le 2bsR^2)\, ds. \tag{1.20}$$

The assumption that μ be Gaussian in Proposition 1 can be relaxed if we merely ask for (1.19) to hold. The following proposition reflects this and is an improvement of Theorem 3 in [9], which was obtained under the assumption that (1.5) holds for all $t > 0$.

PROPOSITION 2. *Let B be a real separable Hilbert space. Assume $D = \{x \in B: \|x-a\| < R\}$ and* (1.4) *holds. If* (A2) *holds, and for all $f \in B^*$* (A3) *holds, then (D, μ) has a unique dominating point a_0 and there exists $c \in (0, 1)$ such that for n sufficiently large*

$$c < n^{1/2} P(S_n \in nD) \exp\{n\lambda(a_0)\} < 1/c. \tag{1.21}$$

REMARKS. 1. For results applicable in general Banach spaces, the analogue of Theorem 2 in [9] can now be proved under the less restrictive moment condition in (A2). For example, if we assume the conditions of part (III) of Theorem 1, then both the upper bound of (1.9) and the lower bound of (1.10) in [9] hold. Unfortunately, an algebra error between (4.11) and (4.13) of [9], page 539, leaves (1.11) in [9] incorrect. What can be shown under the argument given there is that for all $\varepsilon > 0$ and n sufficiently large,

$$P(S_n \in nD) \ge n^{(-1/2+\varepsilon)} \exp\{-n\lambda(a_0)\}.$$

Furthermore, the constant in the upper bound can be made more explicit. That is, an application of the Berry–Esseen theorem, and a standard upper bound for the constant in that theorem, easily implies that $J_n \le Cn^{-1/2}$ in a general Banach space, provided Z and σ_g^2 are as in Proposition 1 and $C = (4E[|g(Z-a_0)|^3] + \sigma_g^2)/\sigma_g^3$. Hence C depends on the second and third absolute moments of $g(Z-a_0)$, but is independent of the convex set D and the Banach space B.

2. Assume (A1), (A2) and that (A3) holds for all $f \in B^*$. Then we have the weak large deviation principle as described in [5] with rate function $\lambda(x)$ as in (1.2). This follows from standard proofs for the upper bound for compact sets in B (even weakly compact sets) by using Theorems 2.3 and 3.2 of [1] applied

to open half-spaces. The lower bound follows for open sets G by showing

$$\liminf_n \frac{1}{n} \log P(S_n \in nG) \geq -\lambda(x) \tag{1.22}$$

for all $x \in G$. Of course, to prove (1.22), simply take an open ball about x entirely within G and again apply Theorems 2.3 and 3.2 of [1]. In fact, the weak large deviation principle holds with less restrictive conditions, and [5], page 278, mentions this in connection with the generalization of Cramér's theorem to the vector space setting.

We conclude this introduction by mentioning an example in the sequence spaces l_p, $1 < p < \infty$. The assumptions (A1), (A2) and (A3) for all $f \in B^*$ hold for this example. However, the related rate function does not have strongly compact level sets, and the exponential integrability in (A2), fails for large $t > 0$. Hence for partial sums of i.i.d. copies of this example, the Donsker–varadhan theory fails to apply and the Bahadur–Zabell results yield logarithmic estimates for open convex sets. However, the extension of Theorem 2 in [9] mentioned in the first remark immediately above yields the upper bound (1.9) and the lower bound (1.10) in [9] for all $p \in (1, \infty)$. If $p = 2$, then Proposition 2 above applies, and (1.21) holds. Hence using the results obtained here, one gets rather detailed estimates of these probabilities.

Let $\{e_k\}$ denote the canonical basis for l_p and assume $\{\xi_k\}$ are independent random variables with $P(\xi_k = \pm \log(k+2)) = p_k$ and $P(\xi_k = 0) = 1 - 2p_k$ where $0 < 2p_k < 1$ for all k and there exists a constant $c \in (0, 1)$ such that $c < k^3 p_k < 1/c$ for all $k \geq 1$. Set

$$X = \sum_{k=1}^{\infty} \xi_k e_k. \tag{1.23}$$

Then some calculation shows X satisfies the claims indicated above.

2. Some useful lemmas. The proof of part (I) of Theorem 1 follows from the next lemma.

LEMMA 2.1. *Let $\lambda(\cdot)$ have weakly sequentially compact level sets and assume $\lambda(x) = 0$ iff $x = m$. If D is an open convex subset of B such that $m \notin D$ and $D \cap \mathrm{dom}(\lambda) \neq \emptyset$, then (D, μ) has a predominating point a_0. Furthermore, $\lambda(x) > \lambda(a_0)$ for all $x \in D$, and if $\lambda(\cdot)$ is strictly convex on* dom(λ)*, then (D, μ) has a unique predominating point.*

PROOF. Since D is convex, its closure in the weak topology and the norm topology are identical. Hence when we write $\overline{D}$, it is unambiguous, but we are always assuming D to be open in the norm topology. The boundary of D is also with respect to the norm topology.

Since $\inf_{x\in\overline{D}}\lambda(x) \le \inf_{x\in D}\lambda(x)$, we first show the reverse inequality. Take $y \in \partial D$, $\lambda(y) < \infty$, and suppose $x_0 \in D$, $\lambda(x_0) < \infty$. Then $[x_0, y) = \{ty + (1-t)x_0: 0 \le t < 1\}$ is a subset of D ([13], page 38), and since λ is convex on $\operatorname{dom}(\lambda)$,

$$\lambda(ty + (1-t)x_0) \le t\lambda(y) + (1-t)\lambda(x_0) < \infty.$$

Thus, since $ty + (1-t)x_0 \in D$,

$$\inf_{x\in D}\lambda(x) \le \liminf_{t\uparrow 1}\lambda(ty + (1-t)x_0) \le \lambda(y),$$

and since $y \in \partial D$ is arbitrary this implies

$$\inf_{x\in D}\lambda(x) = \inf_{x\in\overline{D}}\lambda(x).$$

Hence it now suffices to show there exists $a_0 \in \partial D$ such that $\lambda(a_0) = \inf_{x\in D}\lambda(x)$.

If $m \in \partial D$, take $a_0 = m$, and since $\lambda(m) = 0$, λ nonnegative, we have (1.9)(i). If $m \notin \overline{D}$ we next show that if $x \in (m, x_0)$ where $x_0 \in D$, $x_0 \neq m$ and $\lambda(x_0) < \infty$, then $0 < \lambda(x) < \lambda(x_0)$. This follows since $x \in (m, x_0)$ implies that for some $s \in (0, 1)$, $x = sx_0 + (1-s)m$, and hence by convexity of λ we have

$$\lambda(x) \le s\lambda(x_0) + (1-s)\lambda(m) = s\lambda(x_0) < \lambda(x_0)$$

because $0 < s < 1$ and $\lambda(x_0) > 0$. That $\lambda(x_0) > 0$ and $\lambda(x) > 0$ follows since $\lambda(y) = 0$ iff $y = m$.

Next we observe that if $\inf_{x\in D}\lambda(x) < \infty$ and λ is assumed to have weakly sequentially compact level sets, then there exists a sequence $\{x_j\} \subseteq D \cap \operatorname{dom}(\lambda)$ such that

$$\lim_{j\to\infty} x_j = x_0 \quad\text{and}\quad \lim_{j\to\infty}\lambda(x_j) = \inf_{x\in D}\lambda(x) < \infty. \tag{2.1}$$

Of course, the limit in (2.1) is a weak limit. Hence $x_0 \in \overline{D}$, and since λ is weakly lower semicontinuous (it is the sup of weakly continuous functions), we also have

$$\lambda(x_0) \le \liminf_{j\to\infty}\lambda(x_j) < \infty. \tag{2.2}$$

Thus if we consider the situation $m \notin \overline{D}$, we see $x_0 \neq m$ and $\lambda(x_0) > 0$. Furthermore,

$$0 < \inf_{x\in D}\lambda(x), \qquad x_0 \in \operatorname{dom}(\lambda), \tag{2.3}$$

and λ is strictly nondecreasing on the ray $[m, x_0]$. If $x_0 \in D$, then D open and our previous observation implies $\inf_{x\in D}\lambda(x) < \lambda(x_0)$. Hence $x_0 \in \partial D$ and we take $a_0 = x_0$ to satisfy (1.9)(i).

Now take $x_1 \in D$ and assume $\lambda(x_1) = \lambda(a_0)$. Since $x_1 \in D$, and D is open there exists $s_0 > 0$ such that $0 < s < s_0$ implies $sm + (1-s)x_1 \in D$. Thus $\lambda(sm + (1-s)x_1) \le s\lambda(m) + (1-s)\lambda(x_1) \le (1-s)\lambda(a_0) < \lambda(a_0)$ for $0 < s < s_0$, which contradicts $\lambda(a_0) = \inf_{x\in D}\lambda(x)$. Thus $\lambda(x) > \lambda(a_0)$ for all $x \in D$.

Finally, if $a_0, a_1 \in \partial D$ and $\lambda(a_0) = \lambda(a_1)$ then the ray $[a_0, a_1] \subseteq \overline{D} \cap \mathrm{dom}(\lambda)$ and λ strictly convex on $\mathrm{dom}(\lambda)$ implies $a_0 = a_1$. Hence the lemma is proved. $\square$

LEMMA 2.2. *If $m = \int_B x\, d\mu(x)$ exists and $\hat{\mu}(f) < \infty$ for all f such that $\|f\|_{B^*} \le \varepsilon$ for some $\varepsilon > 0$, then $\lambda(x) = 0$ iff $m = x$.*

The proof follows easily from the proof of Lemma 6.1(ii) in [2].

Now we turn to some further notation. Observe that since $\{x\colon \lambda(x) \le \lambda(a_0)\}$ is a weakly sequentially compact convex subset of B with a_0 satisfying (1.9)(i), and D is an open convex subset of B with $\{x\colon \lambda(x) \le \lambda(a_0)\} \cap D = \varnothing$, then the Hahn–Banach theorem in [8], page 41, allows us to choose $f \in B^*$ such that

$$\sup_{\{x\colon \lambda(x) \le \lambda(a_0)\}} f(x) = f(a_0) = \inf_{x \in D} f(x) < f(x) \qquad \forall\ x \in D. \tag{2.4}$$

Let ψ be defined on B^* by

$$\psi(g) = \int_B x e^{g(x)}\, d\mu(x) / \hat{\mu}(g), \tag{2.5}$$

whenever the integral exists as a Bochner integral. Then, if f is as in (2.4), we define

$$b_t = \psi(tf) \tag{2.6}$$

for all $t \in (-\infty, \infty)$ where $\psi(tf)$ exists (as a Bochner integral).

LEMMA 2.3. *Assume* (A1) *and* (A2). *Let $f \in B^*$ satisfy* (2.4), *assume D is an open convex subset of B satisfying* (1.4), *and assume a_0 is as in Lemma* 2.1. *Then the following hold:*

(2.7) (i) For all t such that b_t exists as a Bochner integral, we have

$$\lambda(b_t) = tf(b_t) - \log \hat{\mu}(tf).$$

(2.7) (ii) If μ^f denotes the μ distribution of f on $(-\infty, \infty)$ with rate function λ_{μ^f}, then for all t such that b_t exists as a Bochner integral,

$$\lambda_{\mu^f}(f(b_t)) = \lambda(b_t).$$

(2.7) (iii) $\mathrm{int}(\mathrm{dom}(\lambda_{\mu^f})) \cap (f(a_0), \infty) \ne \varnothing$.

PROOF. To prove (i), assume b_t exists as indicated, and set $dv_{b_t} = e^{tf(\cdot)}\, d\mu / \hat{\mu}(tf)$. Then by Theorem 3.3b of [1] we have

$$\lambda(x) \le \inf_{v \in \mathscr{M}_x} k(v|\mu), \tag{2.8}$$

where

$$\mathscr{M}_x = \{v\colon v \text{ a probability measure, } \textstyle\int_B y\, dv(y) = x\} \tag{2.9}$$

and

$$k(v|\mu) = \begin{cases} \int_B \log \dfrac{dv}{d\mu} dv, & \text{if } v \ll \mu, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.10}$$

Since b_t exists as a Bochner integral, $v_{b_t} \ll \mu$ and (2.8) imply

$$\begin{aligned} \lambda(b_t) &\le k(v_{b_t}|\mu) \\ &= \int_B [tf(x) - \log \hat{\mu}(tf)]\, dv_{b_t}(x) \\ &= tf(b_t) - \log \hat{\mu}(tf). \end{aligned} \tag{2.11}$$

However,

$$\lambda(b_t) = \sup_{g \in B^*} [g(b_t) - \log \hat{\mu}(g)],$$

so setting $g = tf$ we have

$$\lambda(b_t) \ge tf(b_t) - \log \hat{\mu}(tf). \tag{2.12}$$

Combining (2.11) and (2.12) we thus have (2.7)(i).

Applying f to $b_t = \psi(tf)$, it follows from the previous argument for (2.7)(i) that for all t such that b_t exists as a Bochner integral and $dv_{f(b_t)} = (e^{tu}/\mu^f(t))d\mu$, we have

$$\begin{aligned} \lambda_{(\mu^f)}(f(b_t)) &= k(v_{f(b_t)}|\mu^f) \\ &= t\int_{\mathbb{R}} u\, dv_{f(b_t)}(u) - \log \widehat{\mu^f}(t) \\ &= tf(b_t) - \log \widehat{\mu^f}(t). \end{aligned}$$

Thus (2.7)(ii) holds since $\widehat{\mu^f}(t) = \hat{\mu}(tf)$.

To verify (2.7)(iii) first observe that Theorem 2.4a of [1] implies $\overline{\mathrm{co}(S)} \supseteq \overline{\mathrm{dom}(\lambda)}$. Hence since D is open with $D \cap \mathrm{dom}(\lambda) \ne \varnothing$, we have $\mu(\widetilde{D}) > 0$, where $\widetilde{D} = \{x\colon f(x) > f(a_0)\}$. Thus f satisfying (2.4) implies $\mu^f(f(a_0), \infty) > 0$. Hence

$$\overline{\mathrm{co}(\mathrm{supp}(\mu^f))} \cap (f(a_0), \infty) \ne \phi. \tag{2.13}$$

Now $m \notin \overline{D}$, implies $f(m) < f(a_0)$ and hence

$$\mathrm{int}\big(\overline{\mathrm{co}(\mathrm{supp}(\mu^f))}\big) \ne \phi, \tag{2.14}$$

and (2.13) and (2.14) combine to imply

$$\mathrm{int}\big(\overline{\mathrm{co}(\mathrm{supp}(\mu^f))}\big) \cap (f(a_0), \infty) \ne \phi. \tag{2.15}$$

Applying Theorem 2.4c of [1] and (2.15) thus implies (2.7)(iii). Hence the lemma is proved. □

LEMMA 2.4. *Assume* (A1), (A2) *and that* $f \in B^*$ *satisfies* (2.4), *with* a_0 *and* D *as in Lemma* 2.1. *Then*

$$\lambda(a_0) = \lambda_{\mu^f}(f(a_0)). \tag{2.16}$$

PROOF. Let $\rho_{\mu^f}(s) = \inf\{\lambda(x)\colon f(x) = s\}$. Then $\rho_{\mu^f}(\cdot)$ is the pull-back entropy for μ^f as in [1]. Furthermore, since $f\colon B \to \mathbb{R}$, Theorem 5.2f of [1] implies

$$\rho_{\mu^f}(s) = \lambda_{\mu^f}(s) \tag{2.17}$$

for all $s \in \mathbb{R}$. Now, trivially,

$$\inf\{\lambda(x)\colon f(x) = f(a_0)\} \le \lambda(a_0), \tag{2.18}$$

and by the argument in Lemma 2.1,

$$\inf\{\lambda(x)\colon f(x) > f(a_0)\} = \inf\{\lambda(x)\colon f(x) \ge f(a_0)\}. \tag{2.19}$$

Thus (2.4) implies

$$\inf\{\lambda(x)\colon f(x) > f(a_0)\} = \lambda(a_0). \tag{2.20}$$

Combining (2.18–2.20), we have

$$\inf\{\lambda(x)\colon f(x) = f(a_0)\} = \lambda(a_0),$$

and hence (2.17) with $s = f(a_0)$ implies

$$\lambda(a_0) = \rho_{\mu^f}(f(a_0)) = \lambda_{\mu^f}(f(a_0)).$$

Hence (2.16) holds and the lemma is proved. □

Our next lemma is for measures on the real line. It will later be applied to the measures μ^f, $f \in B^*$, but to simplify notation we will write only μ in the lemma. This lemma, in one form or other, is known, but for lack of a reference in the form we need, we include the details.

LEMMA 2.5. *Let* μ *be a measure on* $\mathbb{R}$ *with* $t^+ = \sup\{t\colon \hat\mu(t) < \infty\} > 0$, $t^- = \inf\{t\colon \hat\mu(t) < \infty\} < 0$, *and let* $\lambda(\cdot)$ *denote the rate function*

$$\lambda(s) = \sup_{t\in\mathbb{R}}[st - \log\hat\mu(t)]. \tag{2.21}$$

Furthermore, assume $\overline{\mathrm{co}(S)}$ *has nonempty interior given by*

$$\mathrm{int}(\overline{\mathrm{co}(S)}) = (a, b), \tag{2.22}$$

and that

$$\lim_{t\nearrow t^+}\frac{\hat\mu'(t)}{\hat\mu(t)} \ge b. \tag{2.23}$$

Let

$$\psi_1(u) = \int_{\mathbb{R}} xe^{ux}\,d\mu(x)/\hat\mu(u) \tag{2.24}$$

provided $xe^{ux} \in L^1(\mu, \mathbb{R})$*, and define*

$$D_{\psi_1} = \{u\colon xe^{ux} \in L^1(\mu, \mathbb{R})\}. \tag{2.25}$$

Then

$$\psi_1(D_{\psi_1}) \supseteq (m, b), \tag{2.26}$$

where $m = \int_{\mathbb{R}} x\, d\mu(x)$*. Furthermore, if* $m < s < b$*, there exists a unique* $t = t_s \in (0, t^+)$ *such that* $\psi_1(t_s) = s$*.*

PROOF. Since $t^+ > 0$ and $t^- < 0$, $m = \int_{\mathbb{R}} x\, d\mu(x)$ exists. Furthermore, (2.22) and the definition of integral then implies $m \in [a, b]$. However, since $(a, b) \neq \phi$ it follows that m is actually in the open interval (a, b). We also observe that if $0 < qu < t^+$, then Hölder's inequality implies

$$\int_{\mathbb{R}} |xe^{ux}|\, d\mu(x) \leq \left(\int_{\mathbb{R}} |x|^p\, d\mu(x)\right)^{1/p} \left(\int_{\mathbb{R}} e^{qux}\, d\mu(x)\right)^{1/q} < \infty. \tag{2.27}$$

In (2.27), $\int_{\mathbb{R}} |x|^p\, d\mu(x) < \infty$ for all $p > 0$, since $t^+ > 0$ and $t^- < 0$. Hence taking $q > 1$ close to one and $1/p + 1/q = 1$, (2.27) implies

$$D_{\psi_1} \supseteq (0, t^+). \tag{2.28}$$

Applying Theorem 2.4c of [1] we have

$$\mathrm{int}(\mathrm{dom}(\lambda)) = (a, b), \tag{2.29}$$

so take $s \in (m, b)$. Now $\lambda(\cdot)$ is finite, nonnegative, convex, and continuous on (m, b), and since $\log \hat{\mu}(t) > tm$, for $s \geq m$ we have that

$$\begin{aligned} \lambda(s) &= \sup_{t \in \mathbb{R}} [st - \log \hat{\mu}(t)] \\ &= \sup_{t \geq 0} [st - \log \hat{\mu}(t)] \\ &= \sup_{0 \leq t \leq t^+} [st - \log \hat{\mu}(t)]. \end{aligned} \tag{2.30}$$

Letting

$$F(t) = \log \hat{\mu}(t),$$

we have F twice continuously differentiable on $(0, t^+)$, and a simple application of the Cauchy–Schwarz inequality implies $F''(t) > 0$ on $(0, t^+)$ with

$$\lim_{t \uparrow t^+} F'(t) = \lim_{t \uparrow t^+} \frac{\hat{\mu}'(t)}{\hat{\mu}(t)} \geq b > s$$

and

$$\lim_{t \downarrow 0} F'(t) = m.$$

Since $s \in (m, b)$, the continuity and strictly increasing nature of F' implies there exists a unique $t_s \in (0, t^+)$ such that

$$F'(t_s) = \frac{\mu'(t_s)}{\mu(t_s)} = s.$$

Thus for $m < s < b$, $G(t) = st - \log \hat{\mu}(t)$ has a finite supremum on $(0, t^+)$ [namely $\lambda(s)$], and for each $s \in (m, b)$,

$$G'(t) = s - \frac{\hat{\mu}'(t)}{\hat{\mu}(t)}$$

has a unique zero at t_s. Since $-G''(t) = F''(t) > 0$ on $(0, t^+)$, this implies that for $s \in (m, b)$ there exists a unique $t_s \in (0, t^+)$ such that

$$\frac{\hat{\mu}'(t_s)}{\hat{\mu}(t_s)} = s. \tag{2.31}$$

For $0, t < t^+$, we have

$$\psi_1(t) = \frac{\hat{\mu}'(t)}{\hat{\mu}(t)},$$

and since $m < s < b$ is arbitrary, (2.31) implies $\psi_1(D_\psi) \supseteq (m, b)$. Hence the lemma is proved. □

LEMMA 2.6. *Assume* (A1), (A2) *and suppose D is an open half-space in B, say $D = \{x\colon f(x) > c\}$. Furthermore, assume D satisfies* (1.4), *and that $\widehat{\mu^f}$ satisfies* (A3). *Then (D, μ) has a dominating point d_0 given by the Bochner integral*

$$d_0 = \int_B x e^{t_0 f(x)}\, d\mu(x)/\hat{\mu}(t_0 f) \tag{2.32}$$

for some $t_0 \in (0, t_f^+)$.

REMARK. If $f(m) = c$, then $m \in \overline{D}$ and we would take $t_0 = 0$ in (2.32) so that $d_0 = m$.

PROOF. Since $f(m) < c$ and $D \cap \mathrm{dom}(\lambda) \neq \phi$ it follows that μ^f is nondegenerate with $\mu^f((c, \infty)) > 0$. Furthermore, using the notation in (1.14), (1.15) and (1.16), (A2) implies $t_f^+ > 0$ and $t_f^- < 0$. In addition, since (A3) holds, we are able to apply Lemma 2.5 to μ^f.

Since $\mathrm{int}(\overline{\mathrm{co}(\mathrm{supp}\, \mu^f)}) = (a_f, b_f)$ and $\mu^f((c, \infty)) > 0$, we have $b_f > c$. Thus Lemma 2.5 implies there exists a unique $t_0 \in (0, t_f^+)$,

$$\int_B f(x) \exp(t_0 f(x))\, d\mu(x)/\hat{\mu}(t_0 f) = c. \tag{2.33}$$

Hence, if $\int_B x \exp(t_0 f(x))\, d\mu(x)$ exists as a Bochner integer, then d_0 given as in (2.32) is such that $d_0 \in \partial D$. Now $\int_B \|x\|^p\, d\mu(x) < \infty$ for all $p < \infty$ by (A2), and if $q > 1$ is sufficiently close to 1 we have $qt_0 \in (0, t_f^+)$. Hence

$$\int_B \exp(qt_0 f(x))\, d\mu(x) < \infty,$$

and Hölder's inequality implies

$$\int_B \|x\| \exp(t_0 f(x))\, d\mu(x) \le \left(\int_B \|x\|^p\, d\mu(x) \right)^{1/p} \times \left(\int_B \exp(qt_0 f(x))\, d\mu(x) \right)^{1/q} < \infty.$$

Thus the integral in (2.32) exists as a Bochner integral with $d_0 \in \partial D$.

Hence $d_0 = b_{t_0}$, with b_t as in (2.6), and Lemma 2.3 implies

$$\lambda(d_0) = \lambda_{\mu^f}(f(b_{t_0})) = \lambda_{\mu^f}(f(d_0)) = \lambda_{\mu^f}(c).$$

However, by Lemma 2.4, $\lambda_{\mu^f}(c) = \lambda(a_0)$ for any predominating point of D, so

$$\lambda(d_0) = \inf_{x \in D} \lambda(x) = \inf_{x \in \overline{D}} \lambda(x).$$

Hence d_0 satisfies (1.9)(i), (ii) and (iv) when $g = t_0 f$. Of course, (1.9)(iii) also holds when $g = t_0 f$ by (2.7)(i). Thus d_0 is a dominating point for (D, μ), and the lemma is proved. □

LEMMA 2.7. *Assume* (A1), (A2), $D = \{x\colon f(x) > c\}$ *satisfies* (1.4), *and* (A3) *holds for this particular* f. *Then* (D, μ) *has a unique predominating point, which is also the dominating point for* (D, μ).

PROOF. We consider several steps.

STEP 1. First we observe (D, μ) has a dominating point, and describe some notation used throughout the proof.

The existence of a dominating point d_0 for (D, μ), given as in (2.32), follows immediately from Lemma 2.6. Hence (D, μ) has at least one predominating point, and given the nature of D and (1.4), we have $f(m) < c$ and that $\mathrm{co}(\mathrm{supp}(\mu^f))$ has nonempty interior. Thus assume a_1, a_2 are distinct predominating points of (D, μ) with $a_1 = d_0$. Then Lemma 2.1 implies $a_1, a_2 \in \partial D$ and $f(a_1) = f(a_2) = c$.

Let $e_1 = a_1 - m$, $e_2 = a_2 - a_1$, and define $\pi\colon B \to \mathbb{R}^2$ by

$$\pi(x) = (f(x), h(x)), \tag{2.34}$$

where $h \in B^*$ is such that $h(e_2) = 1$, $h(e_1) = 0$. The linear functional h exists by the Hahn–Banach Theorem as e_1 and e_2 are linearly independent. Therefore, $\pi(a_1) = (c, h(a_1))$ and $\pi(a_2) = (c, h(a_2))$ with $h(e_2) = h(a_2) - h(a_1) = 1$. Thus $\pi(a_1)$ and $\pi(a_2)$ are distinct points in $\mathbb{R}^2$.

It remains to show that (D, μ) has a unique predominating point. Then (D, μ) has unique dominating point d_0.

STEP 2. Now we consider the measure μ^π on $\mathbb{R}^2$ and some of its properties. Since $x \in D$ iff $f(x) > c$ and $\pi(D) = \{(u, v): f(x) = u,\ h(x) = v,\ x \in D\}$, we have $(u, v) \in \pi(D)$ implies $u > c$. Conversely, if $u > c$ and we take $x = ue_1/f(e_1) + ve_2$, then $f(x) > c$, $h(x) = v$ and $\pi(x) = (u, v) \in \pi(D)$. Thus

$$\pi(D) = \{(u, v) \in \mathbb{R}^2: u > c\}, \tag{2.35}$$

and we also have

$$\pi^{-1}(\pi(D)) = D. \tag{2.36}$$

That is, $D \subseteq \pi^{-1}(\pi(D))$ is obvious, and $y \in \pi^{-1}(\pi(x))$ for $x \in D$ implies $\pi(y) = \pi(x)$. Thus $f(y) = f(x) > c$ and $y \in D$. Hence (2.36) holds.

Thus we see $\mu^\pi(\pi(D)) = \mu(D) > 0$, since D satisfies (2.4), and consequently Theorem 2.4c and Theorem 3.2 of [1] imply $\mu(D) > 0$. Furthermore, the support of μ^π is contained in no affine subspace of $\mathbb{R}^2$. This follows since $m \in \overline{\text{co}(S)}$ and $\mu^\pi(\pi(D)) = \mu(D) > 0$ implies that $\overline{\text{co}(\text{supp } \mu^\pi)}$ contains $\pi(m)$ and points in $\pi(D)$. The points $\pi(a_1)$ and $\pi(a_2)$ are also in $\overline{\text{co}(\text{supp } \mu^\pi)}$ since a_1 and a_2 are both in $\text{dom}(\lambda) \subseteq \overline{\text{co}(\text{supp } \mu)}$. Now it is easy to see $\pi(a_1 - m)$ and $\pi(a_1 - a_2)$ are linearly independent, and hence $\overline{\text{co}(\text{supp } \mu^\pi)}$ contains the nondegenerate triangle formed by $\pi(m)$, $\pi(a_1)$ and $\pi(a_2)$. Thus μ^π has support on no affine subspace of $\mathbb{R}^2$, and hence $\overline{\text{co}(\text{supp } \mu^\pi)}$ has a nonempty interior in $\mathbb{R}^2$. Hence by Theorem 2.4c and Theorem 3.2 of [1],

$$\text{int}(\overline{\text{co}(\text{supp } \mu^\pi)}) = \text{int}(\text{dom}(\lambda_{\mu^\pi})), \tag{2.37}$$

where λ_{μ^π} is the rate function for μ^π. In particular, since $\mu^\pi(\pi(D)) > 0$ with $\pi(D)$ open in $\mathbb{R}^2$, and the set $\overline{\text{co}(\text{supp } \mu^\pi)}$ is the closure of its interior, we have from (2.37) that

$$\text{int}(\text{dom}(\lambda_{\mu^\pi})) \cap \pi(D) \neq \phi. \tag{2.38}$$

Since $a_1, a_2 \in \text{dom}(\lambda)$ and λ has weakly sequentially compact level sets by (A1), we see from Theorem 5.3(v), Theorem 3.2 of [1], and the Eberlein–Smulian theorem that $\pi(a_1)$ and $\pi(a_2)$ are in $\text{dom}(\lambda_\mu^\pi)$. That is, the pulled back entropy function of μ mapped via π is defined by

$$\lambda_0(x) = \begin{cases} \inf\{\lambda(y): y \in \pi^{-1}(x)\}, & \text{if } \pi^{-1}(\{x\}) \neq \phi, \\ +\infty, & \text{if } \pi^{-1}(\{x\}) = \phi \end{cases} \tag{2.39}$$

and hence

$$\lambda_0(\pi(a_i)) \leq \lambda(a_i) < \infty \quad \text{for } i = 1, 2. \tag{2.40}$$

By Theorem 5.3v of [1], (A1) implies

$$\lambda_0(\pi(a_i)) = \lambda_{\mu^\pi}(\pi(a_i)), \qquad i = 1, 2. \tag{2.41}$$

Furthermore, since $\pi^{-1}(\pi(D)) = D$ and $\pi^{-1}(\pi(\overline{D})) = \overline{D}$, where $\overline{D}$ is the closure of D in B, we have from the fact a_1 and a_2 are predominating points of D and (2.39) that

$$\lambda_0(\pi(a_i)) \geq \inf_{x\in\overline{D}} \lambda(x) = \lambda(a_i), \qquad i = 1, 2, \tag{2.42}$$

$$\begin{aligned} \inf_{x\in\pi(D)} \lambda_0(x) &= \inf_{x\in\pi(D)} \inf\{\lambda(y)\colon y \in \pi^{-1}(x)\} \\ &= \inf_{y\in D} \lambda(y) \\ &= \lambda(a_i), \qquad i = 1, 2 \end{aligned} \tag{2.43}$$

and

$$\begin{aligned} \inf_{x\in\pi(\overline{D})} \lambda_0(x) &= \inf_{x\in\pi(\overline{D})} \inf\{\lambda(y)\colon y \in \pi^{-1}(x)\} \\ &= \inf_{y\in\overline{D}} \lambda(y) \\ &= \lambda(a_i), \qquad i = 1, 2. \end{aligned} \tag{2.44}$$

Thus (2.40) to (2.44) combine to imply

$$\lambda_{\mu^\pi}(\pi(a_i)) = \lambda(a_i), \qquad i = 1, 2, \tag{2.45}$$

and since $\lambda_0 = \lambda_{\mu^\pi}$ by Theorem 5.3v of [1] when (A1) holds, we see that

$$\lambda_{\mu^\pi}(\pi(a_i)) = \inf_{x\in\pi(D)} \lambda_{\mu^\pi}(x) = \inf_{x\in\pi(\overline{D})} \lambda_{\mu^\pi}(x), \qquad i = 1, 2. \tag{2.46}$$

Hence from the argument in the proof of Lemma 2.1,

$$\pi(a_i) \in \partial(\pi(D)) \cap \mathrm{dom}(\lambda_{\mu^\pi}), \qquad i = 1, 2.$$

Thus λ_{μ^π} is not strictly convex on $\mathrm{dom}(\lambda_{\mu^\pi})$, as it is constant on $(\pi(a_1), \pi(a_2))$ by (2.46).

STEP 3. The lack of strict convexity of λ_{μ^π} on the ray $(\pi(a_1), \pi(a_2)) \subseteq \pi(\overline{D})$ produces a contradiction.

To check this, define $\nu_{(\alpha,\beta)}$ to be the probability on $\mathbb{R}^2$ whose Radon–Nikodym derivative with respect to μ^π is

$$\frac{d\nu_{(\alpha,\beta)}}{d\mu^\pi}(u, v) = \frac{\exp(\alpha u + \beta v)}{\widehat{\mu^\pi}(\alpha, \beta)}. \tag{2.47}$$

Then the support of $\nu_{(\alpha,\beta)}$ is the same as the support of μ^π for all (α, β) such that $\widehat{\mu^\pi}(\alpha, \beta)$ exists. In particular, for such (α, β), $\mathrm{int}(\mathrm{supp}(\nu_{\alpha,\beta})) \neq \emptyset$. Furthermore, we define

$$\psi_{\mu^\pi}(\alpha, \beta) = \int_{\mathbb{R}^2} (u, v)\, d\nu_{(\alpha,\beta)}(u, v) \tag{2.48}$$

for all $(\alpha, \beta) \in \mathbb{R}^2$ such that the integral in (2.48) exists as a Bochner integral. We also recall $\psi(g)$ given as in (2.5).

Now observe that if $g \in B^*$ is of the form $g(x) = \alpha f(x) + \beta h(x)$, we have

$$\hat{\mu}(g) = \int_{\mathbb{R}^2} \exp(\alpha u + \beta v)\, d\mu^\pi(u, v). \tag{2.49}$$

Furthermore,

$$\pi(\psi(g)) = \psi_{\mu^\pi}(\alpha, \beta), \tag{2.50}$$

provided $\psi(g)$ exists as a Bochner integral. Applying Lemma 2.6, $t_0 f \in \operatorname{int}(\operatorname{dom}(\hat{\mu}))$ and $\psi(t_0 f)$ exists. In addition, from the proof of Lemma 2.6, we see that there exists a $\delta > 0$ such that $\|(\alpha, \beta) - (t_0, 0)\|_{\mathbb{R}^2} < \delta$ implies $g = \alpha f + \beta h \in \operatorname{int}(\operatorname{dom}(\hat{\mu}))$ and $\psi(g)$ exists as a Bochner integral.

Letting $D(\cdot)$ denote differentiation and observing $\log \widehat{\mu^\pi}(\cdot)$ is differentiable on $\operatorname{int}(\operatorname{dom}(\widehat{\mu^\pi}))$, we have

$$D(\log \widehat{\mu^\pi})(\alpha, \beta) = \psi_{\mu^\pi}(\alpha, \beta), \tag{2.51}$$

provided $\|(\alpha, \beta) - (t_0, 0)\|_{\mathbb{R}^2} < \delta$. Differentiating once more, we obtain for such (α, β) that

$$D(\psi_{\mu^\pi})(\alpha, \beta) = \begin{bmatrix} b_{11}(\alpha, \beta) & b_{12}(\alpha, \beta) \\ b_{21}(\alpha, \beta) & b_{22}(\alpha, \beta) \end{bmatrix},$$

where

$$\begin{aligned}
b_{11}(\alpha, \beta) &= \frac{\partial}{\partial \alpha} \int_{\mathbb{R}^2} u\, d\nu_{(\alpha, \beta)}(u, v), \\
b_{12}(\alpha, \beta) &= \frac{\partial}{\partial \beta} \int_{\mathbb{R}^2} u\, d\nu_{(\alpha, \beta)}(u, v), \\
b_{21}(\alpha, \beta) &= \frac{\partial}{\partial \alpha} \int_{\mathbb{R}^2} v\, d\nu_{(\alpha, \beta)}(u, v), \\
b_{22}(\alpha, \beta) &= \frac{\partial}{\partial \beta} \int_{\mathbb{R}^2} v\, d\nu_{(\alpha, \beta)}(u, v).
\end{aligned}$$

Computing $b_{ij}(\alpha, \beta)$ for $i, j = 1, 2$ and $\|(\alpha, \beta) - (t_0, 0)\|_{\mathbb{R}^2} < \delta$, we see

$$D(\psi_{\mu^\pi})(\alpha, \beta) = C(\alpha, \beta), \tag{2.52}$$

where $C(\alpha, \beta)$ is the covariance matrix of $\nu_{(\alpha, \beta)}$. Since $\nu_{(\alpha, \beta)}$ has support in no affine subspace of $\mathbb{R}^2$, this implies $C_{(\alpha, \beta)}$ is strictly positive definite, and hence $\det C(\alpha, \beta) \neq 0$ for $\|(\alpha, \beta) - (t_0, 0)\|_{\mathbb{R}^2} < \delta$. Thus by the inverse mapping theorem, the mapping $\psi_{\mu^\pi}(\cdot)$ is open at all such points (α, β). In particular, it is open at $(t_0, 0)$, so for $\varepsilon > 0$ sufficiently small we have

$$\{(x, y)\colon \|(x, y) - \psi_{\mu^\pi}(t_0, 0)\|_{\mathbb{R}^2} < \varepsilon\} \subset \psi_{\mu^\pi}(U), \tag{2.53}$$

where

$$U = \{(\alpha, \beta)\colon \|(\alpha, \beta) - (t_0, 0)\|_{\mathbb{R}^2} < \delta\}.$$

Now (2.47), (2.48) and (2.50) imply

$$\begin{aligned}\psi_{\mu^\pi}(t_0, 0) &= \pi(\psi(t_0 f + 0h))\\ &= \pi(d_0)\\ &= \pi(a_1).\end{aligned}$$

Thus the nondegenerate ray

$$L(d_0, \varepsilon) = (\pi(d_0), \pi(a_2)) \cap \{(x, y)\colon \|(x, y) - \pi(d_0)\|_{\mathbb{R}^2} < \varepsilon\}$$

is in the range of $\psi_{\mu^\pi}(\cdot)$. This is a contradiction as

$$\lambda_{\mu^\pi}(\psi_{\mu^\pi}(\alpha, \beta)) = k(\nu_{(\alpha, \beta)}|\mu^\pi), \tag{2.54}$$

and $k(\nu_{(\alpha, \beta)}|\mu^\pi)$ is strictly convex in $\nu_{(\alpha, \beta)}$. That is,

$$\begin{aligned}k(\nu_{(\alpha, \beta)}|\mu^\pi) &= \int_{\mathbb{R}^2} \log \frac{d\nu_{(\alpha, \beta)}}{d\mu^\pi}\, d\nu_{(\alpha, \beta)}\\ &= \int_{\mathbb{R}^2} ((\alpha u + \beta v) - \log \widehat{\mu^\pi}(\alpha, \beta)) \frac{\exp(\alpha u + \beta v)}{\widehat{\mu^\pi}(\alpha, \beta)}\, d\mu^\pi(u, v) \qquad (2.55)\\ &\le \lambda_{\mu^\pi}(\psi_{\mu^\pi}(\alpha, \beta)).\end{aligned}$$

However, by Theorem 3.3b of [1] we have

$$\lambda_{\mu^\pi}(\psi_{\mu^\pi}(\alpha, \beta)) \le \inf_{\nu \in \mathscr{M}_{(\alpha, \beta)}} k(\nu|\mu^\pi) \tag{2.56}$$

where

$$\mathscr{M}_{(\alpha, \beta)} = \{\nu\colon \nu \text{ a probability measure, } \textstyle\int_{\mathbb{R}^2} y\, d\nu(y) = \psi_{\mu^\pi}(\alpha, \beta)\}.$$

Combining (2.55) and (2.56) yields (2.54), and hence λ_{μ^π} is strictly convex on the nondegenerate ray $L(d_0, \varepsilon)$. This contradicts the fact that λ_{μ^π} is constant on $(\pi(a_1), \pi(a_2))$. Therefore a_1 and a_2 being distinct is impossible, and (D, μ) must have a unique predominating point. This completes the proof of the lemma. □

3. Proof of Theorem 1. Since we assume (1.4) and (A1), part (I) of the theorem follows immediately from Lemma 2.1.

The proof of part (II) is an immediate consequence of Lemma 2.6.

To prove part (III), take $f \in B^*$ satisfying (2.4), and $H = \{x\colon f(x) > f(a_0)\}$, where a_0 is a predominating point of D given by Lemma 2.1. Then $D \subseteq H$, $\lambda(a_0) > 0$, and H also satisfies (1.4) with $m \notin \overline{H}$ since Lemma 2.2 implies $\lambda(x) = 0$ iff $x = m$ when (A2) holds. That is, $\lambda(m) = 0$, but $\{x\colon \lambda(x) \le \lambda(a_0)\} \cap H = \varnothing$ and applying Lemma 2.1 to H we see any predominating point of H, call it a_1, is such that

$$\lambda(a_1) = \inf_{x \in H} \lambda(x) \ge \lambda(a_0).$$

Thus a_0 and a_1 are both predominating points of H, and since $\widehat{\mu^f}$ satisfies (A3) we have from Lemma 2.6 that H also has a dominating point d_0 as given in (2.32).

Since H has a unique predominating point by Lemma 2.7, then $a_1 = a_2 = d_0$. It remains to show that (A3) holds for all $f \in B^*$ is necessary.

Suppose (A3) fails for some $f \in B^*$. Then μ^f must be nondegenerate on $\mathbb{R}$, and there exists $b_1 < b_2$ such that $f(m) < b_1 < b_2 < b_f$ and

$$\lim_{t \uparrow t_f^+} \frac{d}{dt} \log \widehat{\mu^f}(t) < b_1.$$

If $D = \{x: f(x) > b_2\}$ and $a_0 \in \partial D$ is a dominating point, then by (1.9) there exists $g \in B^*$ such that $a_0 = \psi(g)$ and $D \subseteq M$, where $M = \{x \in B: g(x) > g(a_0)\}$. Note that D as above implies (1.4) holds. Since $a_0 \in \partial D \cap \partial M$, we actually have $D = M$ and $g = sf$ for $s > 0$. Thus

$$\begin{aligned} sf(a_0) &= sf(\psi(g)) \\ &= sf(\psi(sf)) \\ &= s \frac{d}{ds} \log \widehat{\mu^f}(s). \end{aligned}$$

Now $f(a_0) = b_2$ and $s > 0$, so the left-hand term above is sb_2, whereas the right-hand term is less than $sb_1 < sb_2$ by our choice of b_1, b_2 and that the derivative of $\log \widehat{\mu^f}(t)$ is increasing on $(0, t_f^+)$. This is a contradiction, and thus this D cannot have a dominating point. Therefore part (III) is proved. □

4. Proof of Proposition 1. First assume G_2 is nondegenerate. Since (D, μ) has a dominating point with $m \notin \overline{D}$, the representation formula given in (1.10) and (1.11) holds by Theorem 2, and $g(\cdot)$ satisfies a strict inequality in (1.9)(iii). Therefore $\sigma_g^2 > 0$, $g(a - a_0) = 1/b > 0$ and (1.20) holds provided we show

$$\lim_n (2\pi n \sigma_g^2)^{1/2} J_n = \int_0^\infty e^{-s} P(\|G_2\|^2 \le 2bsR^2)\, ds. \tag{4.1}$$

Hence if $Y_{1,j} = g(Z_j - a_0) S(g)/\sigma_g^2$ and $Y_{2,j} = Z_j - a_0 - Y_{1,j}$ for $j \ge 1$, then $g(Y_{2,j}) = 0$ for $j \ge 1$,

$$J_n = E\left(\exp\left(-g\left(\sum_{j=1}^n Y_{1,j}\right)\right)\right) I\left(\sum_{j=1}^n Y_{2,j} \in n(D - a_0) - \sum_{j=1}^n Y_{1,j}\right), \tag{4.2}$$

and $\{Y_{1,j}\}$ and $\{Y_{2,j}\}$ are independent sequences of i.i.d. centered Gaussian random vectors. Let $W_n = \sum_{j=1}^n g(Z_j - a_0)/\sqrt{n}$ and $W_{2,n} = \sum_{j=1}^n Y_{2,j}/\sqrt{n}$. Then W_n is centered Gaussian with variance σ_g^2, $\mathscr{L}(W_{2,n}) = \mathscr{L}(G_2)$ and $W_{2,n}$ is independent of W_n. Hence

$$\begin{aligned} &P(W_{2,n} \in \sqrt{n}(D - a_0) - W_n S(g)/\sigma_g^2 | W_n = u\sigma_g) \\ &\quad = P(W_{2,n} \in \sqrt{n}(D - a_0) - u\sigma_g S(g)/\sigma_g^2), \end{aligned}$$

and conditioning on $W_n = u\sigma_g$, we see from (4.2) that

$$J_n = \int_0^\infty \exp(-\sqrt{n}u\sigma_g) P(W_{2,n} \in \sqrt{n}(D - a_0) - u\sigma_g S(g)/\sigma_g^2) \times \exp\{-u^2/2\}\, du/(2\pi)^{1/2}. \tag{4.3}$$

Setting $k = \sqrt{n}$, $s = ku\sigma_g$ and $h(k,s) = P(W_{2,k^2} \in k(D - a_0) - (s/k)S(g)/\sigma_g^2)$ in (4.3) implies

$$\sqrt{2\pi} k\sigma_g J_{k^2} = \int_0^\infty e^{-s} h(k,s) \exp\{-s^2/2k^2\sigma_g^2\}\, ds. \tag{4.4}$$

Now $x_0 = a - a_0$ and $g(x_0) = 1/b$; therefore $bx_0 - S(g)/\sigma_g^2 \in \{x\colon g(x) = 0\}$, $\{x\colon g(x) = 0\}$ is the support of $\mathscr{L}(G_2)$, and $\{x\colon g(x) = 0\}$ is tangent to the sphere $D - a_0 = \{x\colon \|x - x_0\| < R\}$ at the origin. Thus by the Pythagorean theorem, if $g(x) = 0$, then

$$x \in k(D - a_0) - \frac{bs}{k}x_0 \quad \text{iff } \|x\|^2 < (kR)^2 - \left(k - \left(\frac{s}{k}\right)b\right)^2 R^2. \tag{4.5}$$

Since $(kR)^2 - (k - (s/k)b)^2 R^2 = 2sbR^2 - s^2b^2R^2/k^2$, the above implies

$$\begin{aligned} h(k,s) &= P\left(G_2 - (s/k)(bx_0 - S(g)/\sigma_g^2) \in k(D - a_0) - \frac{bs}{k}x_0\right) \\ &= P\left(\|G_2 - (s/k)(bx_0 - S(g)/\sigma_g^2)\|^2 < 2sbR^2 - s^2b^2R^2/k^2\right). \end{aligned}$$

Thus $P(\|G_2\|^2 \le t)$ continuous in t implies

$$\lim_k h(k,s) = P(\|G_2\|^2 \le 2sbR^2). \tag{4.6}$$

Combining (4.4) and (4.6) implies

$$\lim_k k\sigma_g J_{k^2} = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-s} P(\|G_2\|^2 \le 2sbR^2)\, ds. \tag{4.7}$$

Hence Proposition 1 is proved if G_2 is nondegenerate. If $G_2 = 0$ with probability 1 (G_2 has mean zero), then starting with (4.2) the result is even easier since the indicator function in (4.2) yields integration over the interval $[0, 2\sqrt{n}g(x_0)/\sigma_g]$. That is, (4.3) then becomes

$$\sqrt{2\pi} J_n = \int_0^{2\sqrt{n}/(b\sigma_g)} \exp(-\sqrt{n}u\sigma_g) \exp\{-u^2/2\}\, du,$$

and the proposition now follows as before. $\square$

5. Proof of Proposition 2. If (A2) holds, then the level sets of the rate function λ in (A1) are convex, closed, and norm bounded in B. Since B is a Hilbert space, B is reflexive, and hence these level sets are weakly compact. Hence (A1) holds.

Thus part (III) of Theorem 1, and its proof, imply (D, μ) has a unique dominating point with g in B^* given by $t_0 f$, where f satisfies (2.4). Hence Theorem 2 implies the representation in (1.10) holds with J_n as in (1.11). Furthermore, if $Z, Z_1, Z_2, \ldots$ are i.i.d. random vectors as in (1.11), then the proof of Lemma 2.6 implies $E(\|Z\|^3) < \infty$. Hence the proof of Theorem 3 in [9] applies to yield (1.21). Thus Proposition 2 is proved. $\square$

Acknowledgments. The author thanks Professors Alex de Acosta and Peter Ney for some useful discussions related to the preparation of this paper.

REFERENCES

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

[2] DE ACOSTA, A. and KUELBS, J. (1983). Limit theorems for moving averages of independent random variables. *Z. Wahrsch. Verw. Gebiete* **64** 67–123.

[3] DE ACOSTA, A. and NEY, P. (1998). Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Probab*. **26** 1660–1682.

[4] DEMBO, A. and KUELBS, J. (1998). Refined Gibbs conditioning principle for certain infinite dimensional statistics. *Studia Sci. Math. Hungar.* **34** 107–126.

[5] DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviation Techniques and Applications*. Jones and Bartlett, Boston.

[6] DEUSCHEL, J. and STROOCK, D. (1989). *Large Deviations*. Academic Press, Boston.

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

[8] DUNFORD, N. and SCHWARTZ, J. T. (1964). *Linear Operators I.* Interscience, New York.

[9] EINMAHL, U. and KUELBS, J. (1996). Dominating points and large deviations for random vectors. *Probab. Theory Related Fields* **105** 529–543.

[10] NEY, P. (1983). Dominating points and the asymptotics of large deviations for random walks on $\mathbb{R}^d$. *Ann. Probab*. **11** 158–167.

[11] NEY, P. (1984). Convexity and large deviations. *Ann. Probab*. **12** 903–906.

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

[13] SCHAEFER, H. H. (1970). *Topological Vector Spaces*. Springer, Berlin.

DEPARTMENT OF MATHEMATICS
480 LINCOLN DRIVE
UNIVERSITY OF WISCONSIN
MADISON
WISCONSIN 53706-1388
E-MAIL: kuelbs@math.wisc.edu