Journal of Convex Analysis
Volume 2 (1995), No.1/2, 229–240

Polyhedral Approximation of Convex Sets With an Application to Large Deviation Probability Theory

Peter E. Ney[1]
University of Wisconsin, Dept. of Mathematics, 480 Lincoln Drive, Madison, WI 53706, USA.

Stephen M. Robinson[2,*]
University of Wisconsin, Dept. of Industrial Engineering, 1513 University Avenue, Madison, WI 53706-1572, USA.
e-mail: smr@cs.wisc.edu

Received 8 June 1994
Revised manuscript received 6 January 1995

Dedicated to R. T. Rockafellar on his 60th Birthday

We extend the well known large deviation upper bound for sums of independent, identically distributed random variables in $\mathbb{R}^d$ by weakening the requirement that the rate function have compact level sets (the classical Cramér condition). To do so we establish an apparently new theorem on approximation of closed convex sets by polytopes.

Keywords : Large deviations, Cramér condition, rate function, polyhedral convex sets, polytopes

1991 Mathematics Subject Classification: 60F10, 52A20, 52A25

1. Introduction

Let X_1, X_2, ... be independent random variables taking values in $\mathbb{R}^d$, $1 \leq d < \infty$, with common probability law $\mu(\,\cdot\,) = P\{X_i \in \,\cdot\,\}$, and expectation $EX_i = m = \int x d\mu(x)$ when it exists (the following discussion does not require existence of an expectation). Let $S_n = \sum_{i=1}^{n} X_i$. The law of large numbers says that $S_n/n \to m$ almost surely.
Large deviation theory is concerned with the behavior of S_n/n away from its central

[1] sponsored by an award from the Alexander von Humboldt Stiftung at the University of Frankfurt a.M.
[2] sponsored by the National Science Foundation under Grant CCR-9109345 and the Air Force Systems Command, USAF, under Grant F49620-93-1-0068. The US Government has certain rights in this material, and is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.
* corresponding author.

ISSN 0944-6532 / $ 2.50 © Heldermann Verlag

tendency. A classical result, due originally to Cramér [1], states that there exists a convex, lower semicontinuous function $I : \mathbb{R}^d \to \mathbb{R}_+$, such that

$$\liminf n^{-1} \log P\{S_n/n \in G\} \geq - \inf_{x \in G} I(x) \tag{1.1}$$

for each open subset G of $\mathbb{R}^d$, and that under some hypotheses

$$\limsup n^{-1} \log P\{S_n/n \in F\} \leq - \inf_{x \in F} I(x) \tag{1.2}$$

for each closed F. $I(x)$ is called the rate function, and $\{S_n\}$ is said to satisfy the large deviation principle (LDP).

Convex duality theory plays an important role in the above results, because great interest attaches to properties of $I(\,\cdot\,)$, and because it can be shown that

$$I(x) = \Lambda^*(x), \tag{1.3}$$

where $\Lambda(\alpha) = \log E \exp\langle \alpha, X_1 \rangle$, $\alpha \in \mathbb{R}^d$, and Λ^* is the convex conjugate of Λ. The function Λ is convex, proper, and (by Fatou's lemma) closed, with $\Lambda(0) = 0$.

The inequality (1.1) holds for all d (with no further hypotheses). Also (1.2) is known to hold for $d = 1$, and it holds for $d > 1$ if F is compact, or is closed and convex. For general closed F (1.2) is known to hold if

$$0 \in \operatorname{int} \operatorname{dom} \Lambda, \tag{C}$$

where $\operatorname{dom} \Lambda$ is the effective domain $\{\alpha \in \mathbb{R}^d \mid \Lambda(\alpha) < +\infty\}$. This condition (known as the Cramér condition) ensures that the level sets of Λ^*,

$$L_a(\Lambda^*) = \{x \mid \Lambda^*(x) \leq a\},$$

are compact, a fact that plays a role in the proofs.

The validity of (1.2) for general closed F, without the condition (C), has been an open question with some interesting history. Slaby [10] produced a counterexample when F was *open,* and Dinwoodie [4] extended this to a counterexample for *closed* F with $d = 3$. Yet there are many interesting cases in which (1.2) holds for closed F, and for general F, without (C). In this paper we give some general sufficient conditions under which this is true. A basic difference between our approach and that usually used in proofs of (1.2) is that the latter tries to approximate F itself by halfspaces or balls, whereas we try to approximate the level sets of Λ^*.

In the process we use some apparently new results on the approximation of convex sets by polyhedral convex sets (Theorems 2.1 and 2.2), and on the separation of the level sets of convex functions by polyhedra (Theorem 2.3). Some of these results make use of the principle of cancellation (subtraction) of convex sets established by Rådström [6], in the support-function form studied definitively by Hörmander [5].

For background on large deviation theory the reader is referred to Varadhan [11], Deuschel and Stroock [3], or Dembo and Zeitouni [2]. For background on convexity the reference is of course Rockafellar [9].

2. Results

Here is an outline of our approach to the problem. Let

$$a = \Lambda^*(F) := \inf_{x \in F} \Lambda^*(x).$$

Assume that a is positive, and *suppose* that for each $\epsilon > 0$ we can find a finite collection of open halfspaces $H_1, \dots, H_N$ such that the intersection

$$P = \cap_{i=1}^N H_i$$

satisfies

$$L_{a-\epsilon} \subset P \subset L_a. \tag{2.1}$$

Then for each i, $\Lambda^*(H_i^c) \geq a - \epsilon$, where the superscript c denotes the complement of a set.

There is also no loss of generality in assuming that $\Lambda^*(H_i^c) \leq a$, for if, *e.g.*, $\Lambda^*(H_1^c) > a$, then L_a would be contained in H_1. We claim that in such a case $P' := \cap_{i=2}^N H_i \subset L_a$, so that H_1 could be discarded. Indeed, if this were not so we could find a point z contained in P' but not in L_a. Let $x \in P$, and consider the line segment $[x, z]$. This segment is a subset of P', and by identifying it with part of the real line we can consider it to be ordered with $x < z$. Now z belongs to the open complement of L_a, so the segment $[x, z]$ crosses the boundary of L_a at a point $w < z$. Then $w \in H_1$, which is open; accordingly, if we take w' close enough to w with $w < w' < z$ then $w' \in H_1$. But then $w' \in P$, contradicting our assumption that $P \subset L_a$. Thus we may assume that for $i = 1, \dots, N$

$$a - \epsilon \leq \Lambda^*(H_i^c) \leq a. \tag{2.2}$$

We also note that $F \subset \operatorname{cl} L_a^c$. To see this, let $x \in F$; if $\Lambda^*(x) > a$ then $x \in L_a^c$, so suppose $\Lambda^*(x) = a$. As $a > 0 = -\Lambda(0) = \inf \Lambda^*$, there is some point $p \in \mathbb{R}^d$ with $\Lambda^*(p) < a$. By prolonging the line segment $[p, x]$ slightly beyond x we obtain a point x' at which, by convexity, $\Lambda^*(x') > a$. Therefore $x' \in L_a^c$, so $x \in \operatorname{cl} L_a^c$ and so $F \subset \operatorname{cl} L_a^c$ as asserted. Now from the fact that $P \subset L_a$ we obtain

$$L_a^c \subset P^c = \cup_{i=1}^N H_i^c.$$

The set on the right is closed, and so we have

$$F \subset \operatorname{cl} L_a^c \subset \cup_{i=1}^N H_i^c.$$

Now we can argue that

$$P\{S_n/n \in F\} \;\leq\; P\{S_n/n \in \cup_{i=1}^N H_i^c\} \;\leq\; \sum_{i=1}^N P\{S_n/n \in H_i^c\}. \tag{2.3}$$

But H_i^c is closed and convex, so from the known fact that (1.2) holds for such sets we deduce that

$$\limsup n^{-1} \log P\{S_n/n \in H_i^c\} \leq -\Lambda^*(H_i^c). \tag{2.4}$$

Applying this to (2.3) we find with some calculation that

$$\limsup n^{-1} \log P\{S_n/n \in F\} \;\leq\; -\inf_{i=1}^{N} \Lambda^*(H_i^c) \;\leq\; -\Lambda^*(F) + \epsilon. \tag{2.5}$$

Since $\epsilon > 0$ was arbitrary, it follows that (1.2) holds for F.
We will determine conditions on the probability measure $\mu(\,\cdot\,)$ and the set F that suffice for (2.1). We start by summarizing the convexity results we will need, then give the large deviation results. Proofs are in Section 3.

2.1. Convexity results

Let us start with an arbitrary closed proper convex function $f : \mathbb{R}^d \to (-\infty, +\infty]$ satisfying $f(0) = 0$. For a convex set C, let $\operatorname{cone} C$ be the cone generated by C (that is, the set of all *positive* scalar multiples of points of C). Let $\operatorname{dom} f = \{\alpha \mid f(\alpha) < +\infty\}$; note that $0 \in \operatorname{dom} f$. We write $\operatorname{rc} C$ for the recession cone of C, defined to be the set

$$\{y \in \mathbb{R}^d \mid x + \lambda y \in C \text{ for each } x \in C \text{ and } \lambda \geq 0\},$$

and I_C and I_C^* for its indicator and support functions respectively.
We denote the Euclidean unit ball in $\mathbb{R}^d$ by $B = \{x \mid \|x\| \leq 1\}$, and the unit sphere by $S = \{x \mid \|x\| = 1\}$. The distance from the point x to the set X is

$$d(x, X) = \inf_{y \in X} \|x - y\|,$$

and for sets R and T in $\mathbb{R}^d$,

$$e(R, T) = \sup_{r \in R} d(r, T) = \inf\{\rho \geq 0 \mid T + \rho B \supset R\}$$

is called the *excess* of R over T.
An earlier version of the following theorem appeared in the unpublished report [7].

Theorem 2.1. *Let C be a nonempty closed subset of $\mathbb{R}^d$. Then the following are equivalent:*
(a) *C is convex, $\operatorname{rc} C$ is polyhedral, and $e[C, \operatorname{rc} C] < +\infty$.*
(b) *There is a polyhedral convex cone Z such that for each positive ϵ there exists a finite collection $\{\, x_1, \ldots, x_k \,\}$ of points of $\mathbb{R}^d$ so that with $Q := \operatorname{conv}\{\, x_1, \ldots, x_k \,\}$ one has $Q + Z \subset C \subset Q + Z + \epsilon B$.*
Further, if (b) *holds then $Z = \operatorname{rc} C$ and the points x_i lie in C.*

If (b) holds, we say that C *can be approximated by a polytope.* Here and later we use the term "polytope" interchangeably with "polyhedral convex set" to denote the intersection of finitely many closed halfspaces. (Thus the set P introduced just prior to (2.1) is *not* a polytope.)
Let f^* denote the convex conjugate of f, and for $0 \leq a < \infty$ write

$$L_a = L_a(f^*) = \{x \mid f^*(x) \leq a\}.$$

Theorem 2.2. (Approximation) *Assume that* $a > 0$ *and that*

$$K := \operatorname{cone} \operatorname{dom} f \text{ is polyhedral} \tag{H1}$$

(that is, K is the intersection of a finite collection of closed halfspaces). Then L_a can be approximated by a polytope.

When $a > 0$, the support function $I^*_{L_a}$ of L_a can be written as

$$I^*_{L_a}(\alpha) = \operatorname{cone}(f + a)(\alpha) := \inf_{\tau > 0} \tau^{-1}[f(\tau\alpha) + a]; \tag{2.6}$$

for details see Theorems 9.7 and 13.5 and Corollary 13.2.1 of [9]. If

$$\text{For some } a > 0 \text{ and each } \alpha \in K \cap S \text{ the infimum in (2.6) is attained,} \tag{H2}$$

we shall say later that f satisfies (H2) at a. We show in Lemma 3.2 below that if f satisfies (H2) at a then in fact it satisfies (H2) at every value in $(0, a]$.

Theorem 2.3. (Separation) *Let* $0 < a < b < +\infty$. *Assume that* (H1) *holds and that* f *satisfies* (H2) *at* b. *Then there is a positive* δ *such that*

$$L_a + \delta B \subset L_b. \tag{2.7}$$

Corollary 2.4. (of Theorems 2.2 and 2.3) *Under the hypotheses of* Theorem 2.3 *there exists a polytope* V *such that*

$$L_a \subset V \subset L_b. \tag{2.8}$$

The hypothesis (H1) is an easily visualized geometric condition, but (H2) might be less easy to visualize. For that reason we state here a condition that is stronger than (H2) but is also more geometric. Its proof is given after Lemma 3.2 in Section 3.

Proposition 2.5. *Let* f *be as above and let* $a > 0$. *If for some positive* δ

$$L_a + \delta B \subset \operatorname{dom} f^*, \tag{2.9}$$

then f *satisfies* (H2) *at* a.

2.2. Large deviation results

Finally, we use the above convexity results to prove (in Section 3 below) the following theorem.

Theorem 2.6. *Let* $\Lambda(\alpha) = \log E \exp\langle \alpha, X_1 \rangle$. *Assume that* Λ *satisfies* (H1) *(i.e., that* $\operatorname{cone} \operatorname{dom} \Lambda$ *is polyhedral) and that* Λ *satisfies* (H2) *at* $\Lambda^*(F)$. *Then the large deviation upper bound* (1.2) *holds for* F.

Note that (H2) was used only to obtain the inclusions (2.7) and (2.8).
To obtain the large deviation result (1.2) for closed F a variety of other sufficient conditions can be formulated (see Section 4 below). However, once the Cramér hypothesis is dropped, we know of no conditions that are specified simply and directly in terms of the measure μ or its generating function $\exp \Lambda(\,\cdot\,)$. The attractiveness of the present result is

that it uses a simple condition on Λ, in terms of a well known quantity from convexity theory.

We also point out that under the Cramér condition (C), $\operatorname{cone}\operatorname{dom}\Lambda = \mathbb{R}^d$ so that (H1) holds for all $d \geq 1$. In Section 4 some examples are given where the probabilistic quantities are identified.

3. Proofs

The proof of Theorem 2.1 is deferred to the end of this section. We start with three lemmas, the first two of which concern the function f of Subsection 2.1. The proof of Proposition 2.5 follows Lemma 3.2.

Lemma 3.1. *Assume* (H1). *Then for some positive ϵ,*

$$\sup\{f(\alpha) \mid \alpha \in K \cap \epsilon B\} =: A_\epsilon < +\infty.$$

Proof. As K is polyhedral by (H1), it has a finite set of generators, say $g_1, \dots, g_r$. As $K = \operatorname{cone}\operatorname{dom} f$ there is no loss of generality in taking each g_i to be in $\operatorname{dom} f$. Define a multifunction Γ from $\mathbb{R}^r$ to $\mathbb{R}^d$ by

$$\Gamma(\lambda) = \begin{cases} \sum_{i=1}^r \lambda_i g_i & \text{if } \lambda_1, \dots, \lambda_r \geq 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

Then $K = \Gamma(\mathbb{R}^r)$. By [8, Proposition 2] the set $K \cap B$ is contained in $\Gamma(M)$ for some bounded subset M of $\mathbb{R}^r$. Accordingly, there is a positive constant ζ such that any $k \in K$ can be written as $k = \sum_{i=1}^r \lambda_i g_i$, with each λ_i in the interval $[0, \zeta\|k\|]$. Define $\epsilon = (r\zeta)^{-1}$. If $k \in K \cap \epsilon B$, then we can write $k = \sum_{i=0}^r \lambda_i g_i$ with $\lambda_0 := 1 - \sum_{i=1}^r \lambda_i \in [0,1]$ and $g_0 = 0$. But then by convexity

$$f(k) \leq \sum_{i=0}^r \lambda_i f(g_i) \leq \max_{i=0}^r f(g_i) =: A_\epsilon,$$

which proves Lemma 3.1. □

Lemma 3.2. *Let $0 < a < b < +\infty$. If f satisfies* (H2) *at b then it also satisfies* (H2) *at a.*

Proof. If K is the origin then (H2) holds vacuously, so let $\alpha \in K \cap S$. By hypothesis there is some positive τ_b such that

$$\operatorname{cone}(f+b)(\alpha) = \tau_b^{-1}[f(\tau_b^{-1}\alpha) + b].$$

For any positive τ we have

$$\tau_b^{-1}[f(\tau_b\alpha)+a]+\tau_b^{-1}(b-a) = \tau_b^{-1}[f(\tau_b\alpha)+b] \leq \tau^{-1}[f(\tau\alpha)+b] = \tau^{-1}[f(\tau\alpha)+a]+\tau^{-1}(b-a),$$

so that

$$\tau^{-1}[f(\tau\alpha)+a] \ \geq \ \tau_b^{-1}[f(\tau_b\alpha)+a] + (\tau_b^{-1} - \tau^{-1})(b-a).$$

Accordingly, in seeking the infimum in the definition of cone $(f+a)(\alpha)$ we may restrict τ to the interval $(0,\tau_b]$. Also, we assumed that f was lower semicontinuous with $f(0)=0$. As $a>0$ we see that

$$\lim_{\tau\to 0+} \tau^{-1}[f(\tau\alpha)+a] = +\infty,$$

and so we may further restrict τ to $[\tau_0,\tau_b]$, where τ_0 is some small positive number. Now lower semicontinuity of f implies that the infimum of $\tau^{-1}[f(\tau\alpha)+a]$ is attained. □

Proof of Proposition 2.5. As before, the claim is vacuously true if K is the origin, so we can fix an element $\alpha \in K\cap S$. Recall that the recession function of f, which we write $\operatorname{rec} f$, is the support function of $\operatorname{dom} f^*$ [9, Theorem 13.3]. As we assumed $f(0)=0$ we have

$$\operatorname{rec} f(\alpha) \;=\; \sup_{\tau>0} \tau^{-1}[f(0+\tau\alpha)-f(0)] \;=\; \sup_{\tau>0}\tau^{-1} f(\tau\alpha),$$

and we note for future reference that this difference quotient is nondecreasing in τ [9, Theorem 23.1].
As cone $(f+a) = I^*_{L_a}$, the hypothesis (2.9) yields

$$\operatorname{cone}(f+a) + \delta\|\cdot\| \;\le\; \operatorname{rec} f.$$

From the definitions of cone $(f+a)$ and $\operatorname{rec} f$, we find that there are positive ρ and σ such that

$$\rho^{-1}[f(\rho\alpha)+a] < \operatorname{cone}(f+a)(\alpha)+\delta/2, \qquad \sigma^{-1} f(\sigma\alpha) > \operatorname{rec} f(\alpha) - \delta/2.$$

From the above inequalities we obtain

$$\rho^{-1}[f(\rho\alpha)+a] < \sigma^{-1} f(\sigma\alpha).$$

As the right-hand side is nondecreasing in σ, we see that $\rho<\sigma$ and that whenever $\tau\ge\sigma$ we have

$$\rho^{-1}[f(\rho\alpha)+a] < \tau^{-1} f(\tau\alpha) < \tau^{-1}[f(\tau\alpha)+a],$$

so in looking for the infimum in the definition of cone $(f+a)(\alpha)$ we need only consider $\tau\in(0,\sigma]$. Now we can finish the proof as we did in Lemma 3.2. □

The following lemma is related to more general results in the literature, but as we want a tailored form of the result and do not need the extra generality, we give a simple direct proof.

Lemma 3.3. *Let C be a closed convex subset of $\mathbb{R}^n$. Then*

$$e[C,\operatorname{rc} C] = \sup\{\, I^*_C(\alpha) \mid\; \alpha\in B\cap(\operatorname{rc} C)^\circ \,\}. \tag{3.1}$$

Proof. Write Z for $\operatorname{rc} C$. This cone is closed because C is closed [9, Theorem 8.2]. Denote by σ the supremum on the right side of (3.1). Choose any $c\in C$, and let z be the closest point in Z to c. Then $c=z+\zeta$ with $z\in Z$, $\zeta\in Z^\circ$, and $\langle z,\zeta\rangle = 0$. Evidently $\|\zeta\| = d[c,Z]$. Suppose for the moment that $\zeta\ne 0$ and write $\phi=\zeta/\|\zeta\|$. Then $\phi\in B\cap Z^\circ$, and

$$d[c,Z] = \|\zeta\| = \langle\phi,\zeta\rangle = \langle\phi,c\rangle \le I^*_C(\phi)\le\sigma,$$

and $d[c,Z] \le \sigma$ also in the case $\zeta = 0$ because then $c \in Z$. Now by taking the supremum over $c \in C$ we find that $e[C,Z] \le \sigma$. For the opposite inequality, let $\alpha \in B \cap Z^\circ$. Let $c \in C$ and let z and ζ be as above. Then

$$\langle \alpha, c\rangle \;=\; \langle \alpha, z+\zeta\rangle \;\le\; \langle \alpha, \zeta\rangle \;\le\; d[c,Z],$$

where we used the Schwarz inequality and the fact that $\|\zeta\| = d[c,Z]$. Now take the supremum over $c \in C$ to obtain $I_C^*(\alpha) \le e[C,Z]$, then the supremum over $\alpha \in B \cap Z^\circ$ to conclude that $\sigma \le e[C,Z]$. □

Proof of Theorem 2.2. We apply Theorem 2.1 to $C = L_a$, a closed convex set. By [9, Theorem 27.1(f)], we have $\operatorname{rc} L_a = K^\circ$. Under (H1) this cone is polyhedral. The equation (2.6) implies that $I_{L_a}^* \le f + a$, and then Lemma 3.1 tells us that $I_{L_a}^*(\alpha)$ has a finite supremum on $K \cap B$ (recall that this function is positively homogeneous). Then from Lemma 3.3 we obtain $e[L_a, K^\circ] < +\infty$, which implies that condition (a) of Theorem 2.1 is satisfied. The conclusion of Theorem 2.2 then follows from (b) of Theorem 2.1. □

Proof of Theorem 2.3. If K is the origin then $f = I_{\{0\}}$ and f^* is identically zero, so the conclusion of the theorem holds. Therefore suppose K is not the origin; choose such a and b, and fix any $\alpha \in K \cap S$. Use the attainment hypothesis to find a positive t_b with

$$\operatorname{cone}(f+b)(\alpha) = t_b^{-1}[f(t_b\alpha) + b].$$

Then

$$\operatorname{cone}(f+a)(\alpha) \;=\; \inf_{t>0} t^{-1}[f(t\alpha)+a] \;\le\; t_b^{-1}[f(t_b\alpha)+a] \;<\; \operatorname{cone}(f+b)(\alpha).$$

Now for fixed $a > 0$ $\operatorname{cone}(f+a)(\alpha)$ is closed [9, Theorem 9.7] and convex (as a function of α); its effective domain is K, which is polyhedral, and therefore it is continuous on K by [9, Theorem 10.2]. If $a < b$, then the difference $\operatorname{cone}(f+b) - \operatorname{cone}(f+a)$ is a continuous, strictly positive function on the compact set $K \cap S$. It must therefore have a positive minimum there, say δ. As $\operatorname{cone}(f+c)$ is positively homogeneous for each c, it follows that for each $\alpha \in K$ one has the inequality

$$\operatorname{cone}(f+a)(\alpha) + \delta\|\alpha\| \;\le\; \operatorname{cone}(f+b)(\alpha).$$

But since $\operatorname{cone}(f+a) = I_{L_a}^*$ and $\delta\|\cdot\| = I_{\delta B}^*$, it follows that $L_a + \delta B \subset L_b$. □

Proof of Corollary 2.4. Note first that by (H2) and Theorem 2.3, for some positive δ

$$L_a + \delta B \subset L_b. \tag{3.2}$$

Theorem 2.2 says we can obtain a polytope V such that

$$V \subset L_b \subset V + \delta B. \tag{3.3}$$

Now (3.2) and (3.3) together imply $L_a + \delta B \subset V + \delta B$, which in turn implies the support function inequality

$$I_{L_a}^* + \delta\|\cdot\| \;\le\; I_V^* + \delta\|\cdot\|.$$

Subtracting $\delta\|\cdot\|$ from both sides we obtain $I^*_{L_a} \leq I^*_V$, and as V is closed this implies that $L_a \subset V$. Combining this with (3.3), we obtain (2.8). □

Proof of Theorem 2.6. Choose any closed subset F of $\mathbb{R}^d$ and let $a = \Lambda^*(F)$. The bound (1.2) holds trivially if $a = 0$, so we may assume $a > 0$. Our hypotheses, together with Corollary 2.4, imply that for any $\eta \in (0, a)$ we can find a polytope V satisfying

$$L_\eta \subset V \subset L_a. \tag{3.4}$$

We show that we then have (2.1), and the argument of Section 2 then completes the proof. Let $\epsilon \in (0, a)$ and choose $\eta \in (a - \epsilon, a)$. By Lemma 3.2, Λ satisfies (H2) at η, and then by Theorem 2.3 we have for some positive γ and for the V given by (3.4) the inclusion

$$L_{a-\epsilon} + \gamma B \subset L_\eta \subset V, \tag{3.5}$$

so that V must have nonempty interior. If we write $V = \cap_{i=1}^N C_i$, where the C_i are closed halfspaces whose corresponding interiors (open halfspaces) we denote by H_i, we have $\operatorname{int} V = \cap_{i=1}^N H_i$ by [9, Theorem 6.5]. Then (3.4) and (3.5) imply

$$L_{a-\epsilon} \subset \operatorname{int} L_\eta \subset \operatorname{int} V \subset V \subset L_a,$$

so that we have (2.1). □

Proof of Theorem 2.1.
(a) implies (b): Write Z for $\operatorname{rc} C$. Lemma 3.3, together with our finiteness assumption on $e[C, Z]$, tells us that the effective domain $\operatorname{dom} I^*_C$ of the support function I^*_C is exactly Z° (we always have $\operatorname{cl} \operatorname{dom} I^*_C = Z^\circ$; the point is that here we can remove the closure symbol). But Z° is polyhedral (because Z was assumed to be), hence locally simplicial. Therefore [9, Theorem 10.2] tells us that I^*_C is continuous on Z°.
Now fix $\epsilon > 0$. For each $c \in C$ define $\Psi_c = \{\, \alpha \in Z^\circ \cap S \mid I^*_C(\alpha) \geq \langle \alpha, c\rangle + \epsilon \,\}$. Recall that for each $\alpha \in \mathbb{R}^n$, $I^*_C(\alpha)$ is the pointwise supremum of the linear functions $\langle \alpha, c\rangle$ for $c \in C$. Therefore the intersection of the Ψ_c over $c \in C$ is empty. But the Ψ_c are compact sets because I^*_C is continuous on Z°, so there is a finite subset $\{\, c_1, \ldots, c_k \,\}$ of C such that the intersection of the Ψ_{c_i} for $i = 1, \ldots, k$ is empty. Therefore, for each $\alpha \in Z^\circ$ there is at least one index i with $I^*_C(\alpha) \leq \langle \alpha, c_i\rangle + \epsilon\|\alpha\|$. Writing $s(\alpha) = \sup_{i=1}^k \langle \alpha, c_i\rangle$, we have $I^*_C \leq s + \epsilon\|\cdot\| + I_{Z^\circ}$; note that we have used here the fact that the effective domain of I^*_C is exactly Z°. But s is the support function of the set $Q := \operatorname{conv}\{\, c_1, \ldots, c_k \,\}$, $\epsilon\|\cdot\|$ is the support function of ϵB, and $I_{Z^\circ} = I^*_Z$ is the support function of Z. Writing $P = Q + Z$, we have $C \subset P + \epsilon B$; note that the sum on the right is closed because $Q + \epsilon B$ is compact and Z is closed. But $P \subset C$ because $Q \subset C$ and $Z = \operatorname{rc} C$.
(b) implies (a): Let $\epsilon > 0$ and suppose that Q and Z are as stated in (b). Write $P = Q + Z$; then as $P + \epsilon B$ is closed and convex we have $P \subset C \subset \operatorname{cl} \operatorname{conv} C \subset P + \epsilon B$; accordingly, $e[\operatorname{cl} \operatorname{conv} C, C] \leq e[P + \epsilon B, P] = \epsilon$. As ϵ was arbitrary and C is closed, we have $C = \operatorname{cl} \operatorname{conv} C$ and hence C is convex. Again, because the sets involved are closed, the inclusion $P \subset C \subset P + \epsilon B$ implies $\operatorname{rc} P \subset \operatorname{rc} C \subset \operatorname{rc}(P + \epsilon B) = \operatorname{rc} P$, and therefore $\operatorname{rc} C = \operatorname{rc} P = Z$, which is polyhedral. Finally, the triangle inequality for the excess $e[\,\cdot\,,\,\cdot\,]$

implies that $e[C, \mathrm{rc}\, C] \le e[C, P + \epsilon B] + e[P + \epsilon B, \mathrm{rc}\, C]$. The first term on the right is zero; the second is $e[Q + \epsilon B + Z, Z]$, which is obviously finite. This proves (a).
For the final claim, note that we have already shown, under (b), that $Z = \mathrm{rc}\, C$. As $0 \in Z$ and $Q + Z \subset C$ we have $Q \subset C$, so the x_i lie in C. □

4. Examples and remarks

In this section we first give three examples, the first illustrating a situation in which the Cramér condition fails but our hypotheses apply. The second is a situation in which our hypothesis (H2) is satisfied for some but not all a, and some but not all level sets separate, but in which the Cramér condition is satisfied (so that the large deviation upper bound holds). Thus, separation of level sets is not a necessary condition for (1.2). The third example is a situation in which our hypotheses are not satisfied and in which we do not know whether or not the large deviation upper bound holds. We then conclude with some brief remarks on a possible extension of our approach.

Example 4.1. Λ is not finite on a neighborhood of the origin (i.e., the Crámer condition fails), but Theorem 2.6 applies.
A specific case illustrating this is Example 2.2 of Dinwoodie [4]: μ on $\mathbb{R}^2$ is the direct product of a normal and a Cauchy distribution. Here clearly $0 \notin \mathrm{int}\,\mathrm{dom}\,\Lambda$, but Theorem 2.6 holds. More generally, if μ is any measure on $\mathbb{R}^d$ such that the convex hull of its support is all of $\mathbb{R}^d$, then $\mathrm{dom}\,\Lambda^* = \mathbb{R}^d$, so by Proposition 2.5 Λ satisfies (H2) for all a. If in this situation $d = 2$ and the cone generated by $\mathrm{dom}\,\Lambda$ is any closed cone that is not all of $\mathbb{R}^2$ (e.g., the closed negative quadrant) then the Cramér condition fails. However, Theorem 2.6 applies because (H1) holds (any closed convex cone in $\mathbb{R}^2$ is trivially polyhedral). This would be the case, for instance, if μ is the mixture of a Gaussian measure on $\mathbb{R}^2$ with any measure whose support is the positive quadrant and which has "large" tails.

Example 4.2. Λ satisfies (H2) for some but not all a, some but not all level sets separate, but the large deviation upper bound (1.2) still holds.
Consider the simple binomial distribution $P\{X_1 = 1\} = 1 - P\{X_1 = 0\} = 1/3$. In this case $\Lambda(\alpha) = \log(2 + e^{\alpha}) - \log 3$, and

$$\Lambda^*(x) = \begin{cases} x \log x + (1-x)\log(1-x) - (1-x)\log 2 + \log 3 & \text{if } 0 \le x \le 1, \\ +\infty & \text{otherwise.} \end{cases}$$

As $\mathrm{dom}\,\Lambda = \mathbb{R}$, the Cramér condition (C) applies, so (1.2) holds. The effective domain of Λ^* is the interval $[0,1]$; we have $\Lambda^*(0) = \log(3/2)$ and $\Lambda^*(1) = \log 3$, and the graph has a vertical tangent at each of these points.
Let $\xi_b(s) = s^{-1}[\Lambda(s) + b]$. Then

$$\mathrm{cone}\,(\Lambda + b)(\alpha) = \begin{cases} \alpha \inf_{s>0} \xi_b(s) & \text{if } \alpha \ge 0, \\ \alpha \inf_{s<0} \xi_b(s) & \text{if } \alpha < 0. \end{cases}$$

If $b \in (0, \log(3/2))$ then (H2) must hold (use the Proposition), and the level sets of Λ^* separate. If $b = \log(3/2)$ then (H2) does not hold because the infimum of $\xi_b(s)$ for $s < 0$

is not attained; however, the separation condition (2.7) *does* hold. For $b > \log(3/2)$ (H2) cannot hold because the level sets of Λ^* separate on the right (near 1) only.

Example 4.3. The hypotheses of Theorem 2.6 are not satisfied.
Let Z be a Cauchy random variable in $\mathbb{R}$ and let $X = (Z, Z^2) \in \mathbb{R}^2$. Then

$$\operatorname{dom}\Lambda = \{(\alpha_1, \alpha_2) \in \mathbb{R}^2 \mid \alpha_2 < 0 \text{ or } (\alpha_1, \alpha_2) = (0,0)\}.$$

This is not a *closed* cone (hence not polyhedral), so (H1) is violated.
One can show in this case (with a fair amount of calculating) that Corollary 2.4 fails: that is, the level sets of Λ^* *cannot* be separated by polyhedra. We do not know whether the large deviation upper bound holds in this case for all closed sets.
We conclude by sketching a possible extension of the approach we have taken. Note that in this paper the set F in the upper bound (1.2) has been quite arbitrary, *e.g.* not necessarily closed. Suppose we now add the condition that F is closed, and write $F = (F \cap tB) \cup (F \cap tB^c)$, where B is the unit ball, and t is to be thought of as "large." Then

$$P\{S_n/n \in F\} = P\{S_n/n \in F \cap tB\} + P\{S_n/n \in F \cap tB^c\}.$$

Since $F \cap tB$ is compact, (1.2) is known to hold for this set. As $\Lambda^*(F \cap tB) \geq \Lambda^*(F)$, it suffices to prove that (1.2) holds for $F \cap tB^c$ for some finite t. One could attack this problem by separation methods analogous to those used here, modified to take advantage of the fact that the set $F \cap tB^c$ contains only elements far from the origin. However, that is beyond the scope of this paper.

References

[1] H. Cramér: On a new limit theorem in the theory of probability, in Colloquium on the Theory of Probability, Hermann, Paris, 1937.

[2] A. Dembo, O. Zeitouni: Large Deviations Techniques and Applications, Jones and Bartlett, Boston, 1993.

[3] J. D. Deuschel, D. W. Stroock: Large Deviations, Academic Press, Boston, 1989.

[4] I. H. Dinwoodie: A note on the upper bound for i.i.d. large deviations, Annals of Probability 19, 1991, 1732–1736.

[5] L. Hörmander: Sur la fonction d'appui des ensembles convexes dans un espace localement convexe, Arkiv för Matematik 3, 1955, 181–186.

[6] H. Rådström: An embedding theorem for spaces of convex sets, Proc. Amer. Math. Soc. 3, 1952, 165–169.

[7] S. M. Robinson: An embedding theorem for unbounded convex sets, Technical Summary Report 1321, Mathematics Research Center, University of Wisconsin–Madison, Madison, WI, 1973.

[8] S. M. Robinson: Some continuity properties of polyhedral multifunctions, Mathematical Programming Study 14, 1981, 206–214.

[9] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton, NJ, 1970.

[10] M. Slaby: On the upper bound for large deviations for sums of i.i.d. random vectors, Annals of Probability 16, 1988, 978–990.

[11] S. R. S. Varadhan: Large Deviations and Applications, CBMS–NSF Regional Conference Series in Applied Mathematics No. 46, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1984.