The Annals of Probability
1984, Vol. 12, No. 3, 903–906

CONVEXITY AND LARGE DEVIATIONS

BY PETER NEY

University of Wisconsin, Madison

A convexity argument is used to establish a representation formula from which one can derive the asymptotics of large deviations of sums of i.i.d. random variables on $\mathbb{R}^d$. This simplifies a proof in [4] which relied on fixed point and probabilistic arguments.

The purpose of this addendum to [4] is to observe that the theorem of that paper on the existence of a "dominating point" is really a result in convexity theory, having little direct dependence on its probabilistic origin. The role of convex duality in large deviation theory is well known, (see e.g. [1], [2], [3]), but one can take more direct advantage of the properties of conjugate functions than has been done heretofore. This yields a substantial simplification in the "dominating point" construction, and some slight improvement in hypotheses. The fixed point proof in [4] is now replaced by a simple compactness argument, and more importantly, the fact that the infimum of the Cramèr functional is achieved at an *interior point* of its domain is now almost immediate. This technical point, which is essential in deriving the representation formula from which the large deviation asymptotics follow, required considerable space and effort in [4]. Another advantage of the present approach is that due to its lesser dependence on the properties of random walk, it is more suitable for application to other processes.

The following lemma gathers together the facts we will need on convexity, mostly taken from the book of Rockafellar [5]. Some of the material is also available in the excellent survey paper of Azencott [1].

Let $f: R^d \to R$ be a proper, strictly convex function, and $\mathscr{D}(f) = \{x \in R: f(x) < \infty\}$. Write $\mathring{A}$, $\bar{A}$, and ∂A for the interior, closure, and boundary of a set A. Let $\mathscr{K}_a(f) = \{x: f(x) \le a\}$, $(a < \infty)$ denote the level sets of f, and let $f^*(y) = \sup\{\langle x, y\rangle - f(x)\}$, $y \in R^d$, be the convex conjugate of f. ($\langle , \rangle$ is inner product.) Call f *essentially smooth* ([5], page 251) if $\mathring{\mathscr{D}}(f) \neq \emptyset$, f is differentiable in $\mathring{\mathscr{D}}(f)$, and if $\|\nabla f(x_i)\| \to \infty$ whenever $x_i \to x_0$, for $x_i \in \mathring{\mathscr{D}}$, $x_0 \in \partial\mathscr{D}$.

LEMMA. *Assume that*: (I) *f is essentially smooth*, (II) *$\mathscr{K}_a(f^*)$ is compact for all $a < \infty$*, (iii) *B is a convex set with $[B \cap \mathscr{D}(f^*)]^\circ \neq \emptyset$. Then*

(i) (a) *$\nabla f: \mathring{\mathscr{D}}(f) \to \mathring{\mathscr{D}}(f^*)$ is a homeomorphism,*
(b) *$\nabla f^*(y) = (\nabla f)^{-1}(y)$ for all $y \in \mathring{\mathscr{D}}(f^*)$, and*
(c) *$f^*(y) = \langle y, (\nabla f)^{-1}(y)\rangle - f((\nabla f)^{-1}(y))$;*

(ii) *$\text{Inf}\{f^*(y): y \in B\} \equiv F^*(B)$ is achieved at a* (*unique*) *point $y_0 \in \bar{B} \cap \mathring{\mathscr{D}}(f^*)$;*

(iii) *$\langle (y - y_0), (\nabla f)^{-1}(y_0)\rangle \ge 0$ for all $y \in B$. If $(\nabla f)^{-1}(y_0) \neq 0$ then $y_0 \in \partial B$.*

Received March 1983.
AMS 1980 *subject classifications.* Primary 60G10; secondary 60G50.
Key words and phrases. Large deviations, convexity.

PROOF. (i) is a consequence of [5], Theorem 26.5.

(ii) Take $v \in [B \cap \mathscr{D}(f^*)]^\circ$. Then $f^*(v) < \infty$, and $B_v \equiv \mathscr{K}_{f^*(v)}(f^*) \cap \bar{B}$ is compact. Also clearly $F^*(B) = F^*(B_v)$. By definition of F^* there is a sequence $\{y_n\} \subset B_v$ such that $f^*(y_n) \to F^*(B_v)$, a subsequence $\{n'\} \subset \{n\}$, and a $y_0 \in B_v$ such that $y_{n'} \to y_0$. By the lower semicontinuity of f^* (true for all convex conjugates): $\liminf f^*(y_{n'}) = F^*(B) \geq f^*(y_0)$.

Now take any point $w \in [B \cap \mathscr{D}(f^*)]^\circ$, and let $z_\gamma = \gamma w + (1-\gamma)y_0$ for $0 < \gamma < 1$. The essential smoothness of f implies that of f^* ([5] Theorem 26.5), and hence one can infer from Lemma 26.2 of [5] that if $y_0 \in \partial\mathscr{D}(f^*)$, then there exists a sequence $\gamma_n \searrow 0$ such that $f^*(z_{\gamma_n})$ is strictly increasing. On the other hand $f^*(y_0) \leq \inf\{f^*(y); y \in B\} \leq f^*(z_{\gamma_n})$, and hence by the strict convexity of f^* ([5], Corollary 26.4.1) we see that for any $0 < \alpha < 1$

$$f^*(z_{\gamma_n}) \geq \alpha f^*(z_{\gamma_n}) + (1-\alpha)f^*(y_0) > f^*(\alpha z_{\gamma_n} + (1-\alpha)y_0).$$

But we can now choose α so that $\alpha z_{\gamma_n} + (1-\alpha)y_0 = z_{\gamma_{n+1}}$, implying $f^*(z_{\gamma_n}) > f^*(z_{\gamma_{n+1}})$. This is a contradiction, and hence we must have $y_0 \in \mathring{\mathscr{D}}(f^*)$. Thus f^* is also continuous at y_0 and $f^*(y_0) = F^*(B)$.

(iii) We claim that $B_{y_0} = \{y_0\}$ (the one point set). Clearly, $y_0 \in B_{y_0}$. Suppose also $y_1 \in B_{y_0}$ for some $y_1 \neq y_0$. Now $y_1 \in \mathscr{K}_{f^*(y_0)}(f^*)$ implies $f^*(y_1) \leq f^*(y_0)$, and the strict convexity of f^* implies that $f^*(\frac{1}{2}(y_0 + y_1)) < f^*(y_0)$. But since $\frac{1}{2}(y_0 + y_1) \in \bar{B}$, this contradicts $f^*(y_0) = F^*(B)$. Thus $\mathscr{K}_{f^*(y_0)}(f^*) \cap \bar{B} = \{y_0\}$ and hence there exists a separating hyperplane H of $\mathscr{K}_{f^*(y_0)}(f^*)$ and B.

Since $y_0 \in \mathring{\mathscr{D}}(f^*)$, f^* is differentiable at y_0. If $\nabla f^*(y_0) = 0$, then (iii) is trivially satisfied. If $\nabla f^*(y_0) \neq 0$, then $\nabla f^*(y_0)$ is the unique normal to $\mathscr{K}_{f^*(y_0)}(f^*)$ at y_0; and hence the separating hyperplane H must be the unique hyperplane through y_0 and orthogonal to $\nabla f^*(y_0)$. Thus $y_0 \in \partial B$ and (iii) follows. □

We now apply the lemma to the estimation of large deviation probabilities. Let μ be a probability measure on R^d;

$$\varphi(\alpha) = \int e^{\langle \alpha, x\rangle}\mu(dx), \quad \alpha \in R^d; \quad m = \int x\mu(dx);$$

$\mathscr{S}$ = closure of the convex hull of the support of μ; B a Borel set in R^d.

DEFINITION. A point v_B is called a dominating point for (B, μ) if

(I) $$v_B \in \partial B \cap \mathring{\mathscr{S}} \cap \mathscr{R}(\nabla \log \varphi), \quad \text{and}$$

(II) $$B \subset \{x: \langle x, \alpha_B\rangle > \langle v_B, \alpha_B\rangle\}, \quad \alpha_B = (\nabla \log \varphi)^{-1}(v_B).$$

THEOREM. *If $\mathscr{D}(\varphi)$ contains a neighborhood of the origin,* $\log \varphi$ *is essentially smooth, B is convex, $[B \cap \mathscr{S}]^\circ \neq \emptyset$, and $m \notin \bar{B}$, then a unique dominating point for (B, μ) exists.*

PROOF. We apply the lemma with $f = \log \varphi$. The compactness of $\mathscr{K}_a(f^*)$ follows from the convergence of φ in a neighborhood of 0 (Azencott, [1]). By Proposition 9.7 of [1] one can also identify $\mathring{\mathscr{D}}(f^*) = \mathring{\mathscr{S}}$. (These are the only places where any probabilistic arguments enter.) The hypotheses of the lemma are thus satisfied, and the point $y_0 = v_B$ of the lemma is the dominating point. Namely $v_B \in \partial B \cap \mathring{\mathscr{S}}$ follows from (ii) of the lemma; also $v_B \in \mathscr{R}(\nabla \log \varphi)$ (i.e. $\nabla \log \varphi(\alpha) = v_B$ has a solution $\alpha(v_B) \in \mathscr{D}(\log \varphi)$), since by (i) $(\nabla f)^{-1}(v_B) \in \mathscr{D}(\log \varphi)$. This implies (I) in the definition, and (II) is just a translation of (iii) in the lemma. □

One can now easily derive the formula (see [4]):

$$\mu^{*n}(nB) = \rho^n \int_{n(B-v_B)} \exp(-\langle \alpha_B, x\rangle)\tilde{\mu}^{*n}(dx), \tag{1}$$

where

(i) $\rho = \exp(-f^*(v_B)) = \exp(-\langle \alpha_B, v_B\rangle)\varphi(\alpha_B)$
(ii) $\langle \alpha_B, x\rangle \geq 0$ for $x \in B - v_B$, and
(iii) $\tilde{\mu}$ is a probability measure with mean 0, which is a centering of the conjugate distribution of μ. From (1) there follow a variety of estimates of $\mu^{*n}(nB)$, the easiest of which is

$$c_1 \rho^n n^{-d/2} \leq \mu^{*n}(nB) \leq c_1 \rho^n n^{-1/2}, \quad n \geq 0, \tag{2}$$

for some $0 < c_1 \leq c_2 < \infty$. This in turn of course implies the known "logarithmic" limit law ([2], [3])

$$\lim (1/n) \log \mu^{*n}(nB) = \log \rho, \tag{3}$$

but much sharper results can also be obtained (see [4]).

REMARKS ON HYPOTHESES. (i) Instead of assuming that $\log \varphi$ is ess. smooth (in the theorem) we could take it to be closed, since with its strict convexity this also implies $(\log \varphi)^*$ is ess. smooth ([5], 26.3); and this is what is needed in the proof. Either conditions is weaker than "$\mathscr{D}(\varphi)$ open," as assumed in [4].

(ii) The inequality (2) was proved in [4] under the additional hypothesis (*) that μ was either lattice or strongly nonlattice, which was used to obtain an estimate $\tilde{\mu}^{*n}(A) \geq cn^{-d/2}$ for a sufficiently large set A, where $\tilde{\mu}$ is the measure in (1). This inequality is in fact true without (*), as can be seen from a local limit theorem of Stone [6]. This fact, as well as an independent proof of the inequality, was shown to me by H. Carlsson. Thus (3) also follows from the "stronger" estimate (2) without extraneous "lattice" conditions (as was known).

Acknowledgment. I would like to thank Professors H. Carlsson and A. de Acosta for several helpful conversations on this subject. Professor de Acosta has communicated to me another "convexity" proof of the dominating point theorem.

ADDED IN PROOF. A recent paper by R. Ellis in *The Annals of Probability* **12** 1–12 exploits convexity arguments along similar lines to this note.

REFERENCES

[1] AZENCOTT, R. (1980). *Grandes Déviations et Applications. Lecture Notes in Math. 774.* Springer, New York.

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

[3] BARTFAI, P. (1978). Large deviation of the sample mean in Euclidean spaces. *Purdue Univ. Statist. Dept. Mimeo Series No.* 78-13.

[4] NEY, P. (1983). Dominating points and the asymptotics of large deviations for random walk on R^d. *Ann. Probab.* **11** 158–167.

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

[6] STONE, C. (1965). On local and ratio limit theorems. *Fifth Berkeley Symp. on Statist. and Probab.* **2** Part 2, 217–224.

DEPARTMENT OF MATHEMATICS
VAN VLECK HALL
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706