The Annals of Probability
1979, Vol. 7, No. 4, 587–621

LARGE DEVIATIONS OF THE SAMPLE MEAN IN GENERAL VECTOR SPACES[1]

BY R. R. BAHADUR AND S. L. ZABELL

University of Chicago

Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random vectors taking values in a space V, let $\bar{X}_n = (X_1 + \cdots + X_n)/n$, and for $J \subset V$ let $a_n(J) = n^{-1} \log P(\bar{X}_n \in J)$. A powerful theory concerning the existence and value of $\lim_{n\to\infty} a_n(J)$ has been developed by Lanford for the case when V is finite-dimensional and X_1 is bounded. The present paper is both an exposition of Lanford's theory and an extension of it to the general case. A number of examples are considered; these include the cases when X_1 is a Brownian motion or Brownian bridge on the real line, and the case when $\bar{X}_n$ is the empirical distribution function based on the first n values in an i.i.d. sequence of random variables (the Sanov problem).

1. Introduction. Let V be a topological vector space, and let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random vectors taking values in V. For each $n = 1, 2, \cdots$ let

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}. \tag{1.1}$$

Let J be a Borel measurable subset of V, let

$$\mu_n(J) = P(\bar{X}_n \in J) \tag{1.2}$$

and let

$$a_n(J) = n^{-1} \log \mu_n(J). \tag{1.3}$$

We shall say that the set J has entropy $s(J)$ if $\lim_{n\to\infty} a_n(J)$ exists and equals $s(J)$, $-\infty \leqslant s \leqslant 0$. This paper discusses certain methods for determining whether $s(J)$ exists and if so for finding its value.

The estimation of $\mu_n(J)$ was first considered by Cramér (1938) in the case when the X_i are real valued, the common distribution of the X_i has an absolutely continuous component, J is an infinite interval, say $J = [a, \infty)$, and certain other conditions are satisfied. Cramér obtained an asymptotic expansion of $\mu_n(J)$ which shows that, in his case, the interval J has entropy given by

$$s([a, \infty)) = \inf\{-at + \log \phi(t) : t \geqslant 0\} \tag{1.4}$$

where $\phi(t) = E(\exp(tX_1))$ for $-\infty < t < \infty$, $0 < \phi(t) \leqslant \infty$. It was shown by

Received May 11, 1977; revised June 24, 1978.

[1]Support for this research was provided in part by National Science Foundation Grant No. MCS72-04364 A04, and by U.S. Energy Research and Development Administration Contract No. EY-76-S-02-2751.

AMS 1970 *subject classifications.* Primary 60F10; Secondary 62F20, 62G20.

Key words and phrases. Random vectors, large deviations, entropy, Sanov's theorem, exponential family, maximum likelihood.

Chernoff (1952) that, whatever the distribution of X_1, the entropy of any interval $[a, \infty)$ exists and is given by (1.4). The initial work of Cramér and Chernoff has been continued and extended in several directions by Bahadur and Ranga Rao (1960), Sethuraman (1964, 1965), Hoeffding (1965a), Borovkov and Rogozin (1965), Sievers (1975), Bartfai (1978), and many others; cf. Petrov (1975), Ibragimov and Linnik (1971), and Bahadur (1971) for other references.

The following elegant generalization of the Cramér-Chernoff problem was proposed by Sanov (1957). Suppose $Y_1, Y_2, \cdots$ is a sequence of i.i.d. real valued random variables, and for each n let $\hat{F}_n$ be the empirical distribution function when the sample is $(Y_1, \cdots, Y_n)$. Let J be a given set of probability distribution functions on the real line. The Sanov problem is to show that $n^{-1} \log P(\hat{F}_n \in J)$ tends to a particular limit defined in terms of the Kullback-Leibler information numbers. An exposition of the relation between Chernoff's theorem and Sanov's problem is given in Bahadur (1971). Sanov's work has been clarified and extended by Hoadley (1967); see also Hoeffding (1965b), Sethuraman (1970), Borovkov (1967), Stone (1974), Sievers (1976), and Groeneboom, Oosterhoff and Ruymgaart (1979). A useful large deviation result for rank statistics was obtained in the spirit of Sanov's work by Woodworth (1970); cf. also Ho (1974).

Lanford (1973) has made a penetrating attack on the problems described in the first paragraph of this section. He proceeds independently of all the work cited above, and uses methods developed by him and Ruelle (1965) in statistical mechanics; indeed, he considers his deep excursion into this problem area merely a digression and source of amusing examples! Lanford treats mainly the case when V is the finite dimensional Euclidean space R^k and X_1 is a bounded random vector. In the following Sections 2 to 5 we present an exposition and generalization of Lanford's theory to the case of a possibly infinite dimensional V and possibly unbounded X_1. Some of our results concerning the general case are closely related to or even identical with certain results of Donsker and Varadhan (1975, 1976). We think that the work just cited and the present one complement each other in various ways. We should add that the cited papers of Lanford, Donsker, and Varadhan are much more extensive in scope than might be suggested by our subsequent references to some of their specific contents.

The present paper seems worthwhile to us for the following reasons. Our generalization of Lanford's theory is not entirely trivial or mechanical, and it includes a wide variety of examples as special cases. In particular, given a sequence $Y_1, Y_2, \cdots$ of i.i.d. real valued random variables, for each i let X_i be the distribution function degenerate at Y_i; then $X_1, X_2, \cdots$ is a sequence of independent and identically distributed random elements in the vector space of functions of bounded variation on the real line, and $\bar{X}_n = (X_1 + \cdots + X_n)/n$ is $\hat{F}_n$. Thus the Sanov problem can be studied in the present framework, and it is so studied in Section 7. Finally, we think that some of the present considerations throw light on exactly what are the real difficulties of the present large deviation problem in general and of the Sanov problem in particular. Our conclusions in this regard are described in subsequent paragraphs of this section and in Section 7.

Now consider the framework of the initial paragraph of this section. Let μ be the probability measure induced on V by the random vector X_1, i.e.,

$$\mu(J) = P(X_1 \in J) \tag{1.5}$$

for Borel sets $J \subset V$; of course μ is μ_1 of (1.2). Assume henceforth that V is a locally convex Hausdorff space and that certain further assumptions (Assumptions 1, 2, and 3 below) also hold. A sufficient condition for these further assumptions to hold is that V be a separable and complete metric space (cf. Lemma 1.1 below).

It is shown in Section 2 that $s(J)$ exists if J is a finite union of open convex sets (Theorem 2.1). Define the point entropy function $s(v)$ in terms of the set entropy function $s(J)$ by

$$s(v) = \inf\{s(J) : v \in J, J \text{ open convex}\} \tag{1.6}$$

for $v \in V$, $-\infty \leqslant s \leqslant 0$. Then $s(v)$ is a concave, proper, and upper semicontinuous function (Theorem 2.2, Corollary 2.1). For any $J \subset V$ let

$$\operatorname{lan}(J) = \sup\{s(v) : v \in J\} \tag{1.7}$$

if $J \neq \varnothing$, and let $\operatorname{lan}(J) = -\infty$ if $J = \varnothing$. Then, one hopes,

$$s(J) = \operatorname{lan}(J) \tag{1.8}$$

for a large class of sets. This hope is based on the fact (cf. Lemma 2.1(b)) that the entropy of a set equals, so to speak, the entropy of its most likely atom. It is shown that (1.8) does hold if J is a finite union of open convex sets (Theorem 2.3). It is also shown that in general (1.8) does not hold for all open sets J (Example 7.1).

In Section 3 we consider some methods of evaluating the point entropy $s(v)$. Let θ denote a continuous linear functional on V, and let V^* be the set of all such functionals. With μ defined by (1.5) let c be the cumulant generating function of μ, i.e.,

$$c(\theta) = \log \textstyle\int_V \exp[\theta(v)]\,\mu(dv) \qquad \text{for} \quad \theta \in V^*, -\infty < c \leqslant \infty. \tag{1.9}$$

Let c^* denote the Fenchel transform of c, i.e.,

$$c^*(v) = \sup\{\theta(v) - c(\theta) : \theta \in V^*\} \qquad \text{for} \quad v \in V. \tag{1.10}$$

Since $\theta(v)$ and $c(\theta)$ vanish if θ is the zero functional, $0 \leqslant c^*(v) \leqslant \infty$ for all v. Let Θ be the set of all θ in V^* for which $c(\theta)$ defined by (1.9) is finite; we shall call Θ the natural parameter space. It is plain that θ may be restricted to Θ in the definition (1.10) of $c^*(v)$. It is shown in Section 3 that

$$s(v) = -c^*(v) \tag{1.11}$$

for every v in V (Theorem 3.2). We defer description of other methods of finding $s(v)$ to Section 3, but note here that perhaps the simplest and most useful method is to deploy the exponential family of measures $\{\mu_\theta : \theta \in \Theta\}$ associated with μ (Corollary 3.3); this method is, however, not always available (Examples 3.1).

Now let J be a Borel measurable subset of V which is not a finite union of open convex sets, and consider the problem of finding asymptotic bounds for $a_n(J)$ defined by (1.2) and (1.3). Let J° denote the interior of J. Then

$$\liminf_{n\to\infty} a_n(J) \geqslant \operatorname{lan}(J^\circ) \tag{1.12}$$

where lan is given by (1.7). To see this, let $v \in J^\circ$. There exists an open convex $A \subset J$ such that $v \in A$. It follows from (1.2), (1.3) that $a_n(J) \geqslant a_n(A)$ for every n; hence $\liminf_{n\to\infty} a_n(J) \geqslant s(A) \geqslant s(v)$ by (1.6); since $v \in J^\circ$ is arbitrary, (1.12) follows. It seems much more difficult to obtain upper bounds, e.g.,

$$\limsup_{n\to\infty} a_n(J) \leqslant \operatorname{lan}(J), \tag{1.13}$$

or even

$$\limsup_{n\to\infty} a_n(J) \leqslant \operatorname{lan}(\bar{J}), \tag{1.14}$$

where $\bar{J}$ is the closure of J. It seems to us that (1.13) or (1.14) require special conditions or special methods in concrete cases. Some general sufficient conditions for (1.13) are that J be compact (Lemma 2.5) or that J be a closed convex set (Lemma 2.6). The point for us is, of course, that if (1.13) or (1.14) holds and J is regular in the sense that $\operatorname{lan}(J^\circ) = \operatorname{lan}(J)$ or $\operatorname{lan}(\bar{J})$, one can conclude that the entropy of J exists and (1.8) holds.

Sometimes it is possible to establish (1.8) in other ways. Suppose for example that $J = K'$ where K is a compact convex set with nonempty interior; here, and subsequently, A' denotes the complement of the set A. With no loss of generality assume that the origin is in the interior of K. Let α be a constant, $0 < \alpha < 1$, and let $\{H\}$ be the set of all hyperplanes H supporting αK. For each H let L denote the open half-space determined by H which does not contain any point of αK. Then $\{L\}$ is an open covering of the boundary of K; let $L_1, \cdots, L_k$ be a subcovering. Then $J \subset \cup_i L_i = M$ say; hence $a_n(J) \leqslant a_n(M)$ for every n; hence $\limsup_{n\to\infty} a_n(J) \leqslant s(M) = \operatorname{lan}(M) \leqslant \operatorname{lan}((\alpha K)')$, since $M \subset (\alpha K)'$. Assuming that $\operatorname{lan}((\alpha K)')$ tends to $\operatorname{lan}(J)$ as $\alpha \to 1$, it follows that (1.13) holds. Since J is open, (1.12) now implies that (1.8) holds. This argument is used by Abrahamson (1965) and Gupta (1972) in certain finite dimensional cases. A related argument is used by Sethuraman (1964) to show that if V is a separable Banach space, if condition (1.15) below holds, if $E(X_1 | \mu) = 0$, and if $J_\varepsilon = \{v : \|v\| \geqslant \varepsilon\}$ where $0 < \varepsilon < \infty$, then $s(J_\varepsilon)$ exists and equals the supremum of the entropies of all closed half-spaces contained in J_ε. It can be shown that here $\operatorname{lan}(J_\varepsilon)$ and $\operatorname{lan}(\{v : \|v\| = \varepsilon\})$ are alternative formulae for $s(J_\varepsilon)$.

Donsker and Varadhan (1975, 1976) have studied asymptotic bounds for $a_n(J)$ when $\{X_n\}$ is a Markov process on a Polish state space. When specialized to the present case of an i.i.d. process, the lower bounds of Donsker and Varadhan are equivalent to (1.12), but their upper bounds provide important complements to the present version of Lanford's theory. For example, Theorem 5.3 of Donsker and

Varadhan (1976) implies that if V is a separable Banach space, and if

$$\int_V \exp[t\|v\|]\,\mu(dv) < \infty \qquad \text{for all} \quad t > 0, \tag{1.15}$$

then (1.13) holds for all closed sets J. It follows hence that if V is a separable Banach space and (1.15) holds then $s(J)$ exists and equals lan(J) for every Borel set J such that lan(J°) = lan($\bar{J}$). We use several results of Donsker and Varadhan (1975, 1976) in subsequent sections.

Our conclusion that it is harder to find adequate asymptotic upper bounds for $a_n(J)$ than to find adequate asymptotic lower bounds is at variance with the superficial appearance of some of the purely probabilistic literature on the subject. The conclusion is, however, in accordance with the well-known fact (cf., e.g., Bahadur (1967)) that in large sample theories of inference it is relatively easy to establish certain universal lower bounds for the rate of convergence of estimates or test-statistics but much harder to establish that these bounds are attainable. The accordance just mentioned is not fortuitous; some of the connections are described and used in Bahadur and Zabell (1979).

A number of examples are presented in terms of Lanford's theory in Sections 6 and 7. Our infinite-dimensional examples are the case when V is a Hilbert space and μ is a Gaussian probability measure (Example 6.3); the case when V is the space of continuous functions on [0, 1] and μ is Wiener measure (Example 6.4); the case when X_1 is a Brownian bridge on [0, 1] (Example 6.5); and the Sanov problem (Section 7). There are some indications that if V is infinite-dimensional, and there are several topologies on V which meet all our requirements, then the point entropy function is independent of which of these topologies is introduced on V (cf. Corollary 5.1). Assuming that this interesting independence does hold in a given case, what topology to introduce on V involves considerations such as the following: a large topology will have a large class of open convex sets (for which (1.8) is known to hold), but it will have a small class of compact sets (for which (1.13) is known to hold), and it will be relatively hard to compute $s(v)$ from (1.10) and (1.11) since V^* will be a large set.

We conclude this section with a formal statement of the present model and of the assumptions required in Sections 2 to 5. Let V be a real vector space of points v, and τ a given topology on V. We assume that, under τ, V is a locally convex (Hausdorff) topological vector space. Let $\mathfrak{B}$ be the σ-algebra of Borel sets of V, and μ a given probability measure on $\mathfrak{B}$.

Let Ω be a space of points ω, $\mathcal{A}$ a σ-algebra of sets of Ω, P a probability measure on $\mathcal{A}$, and $\{X_n : n = 1, 2, \cdots\}$ a sequence of $\mathcal{A} - \mathfrak{B}$-measurable transformations of Ω into V such that, for each n and $B_1, \cdots, B_n$ in $\mathfrak{B}$,

$$P(X_i(\omega) \in B_i \text{ for } i = 1, \cdots, n) = \Pi_{i=1}^n \mu(B_i). \tag{1.16}$$

As is well known, such entities $(\Omega, \mathcal{A}, P)$ and $\{X_n\}$ always exist; these entities are not really essential here but they facilitate certain descriptions, arguments, and verifications.

For $m, n = 1, 2, \cdots$ let $Y_{m,n}(\omega) = \Sigma_{i=m}^{m+n-1} X_i(\omega)/n$.

ASSUMPTION 1. (a) *$Y_{m,n}$ is an $\mathcal{A}-\mathcal{B}$-measurable transformation of Ω into V.* (b) *For all B_1, B_2 in $\mathcal{B}$, $P(Y_{1,m} \in B_1, Y_{m+1,n} \in B_2) = P(Y_{1,m} \in B_1) \cdot P(Y_{m+1,n} \in B_2)$.* (c) *For all B in $\mathcal{B}$, $P(Y_{m+1,n} \in B) = P(Y_{1,n} \in B)$.*

Part (a) of Assumption 1 with $m = 1$ implies (cf. (1.1), (1.2)) that, for each n, μ_n is a well-defined probability measure on $\mathcal{B}$. It is known that in the general case Assumption 1 is not automatically fulfilled.

Let λ be a probability measure on $\mathcal{B}$. We shall say that λ is *regular* if for each open set $B \subset V$, $\lambda(B) = \sup\{\lambda(K) : K \subset B, K \text{ compact}\}$. We shall say that λ is *convex-regular* if for every open convex set $J \subset V$, $\lambda(J) = \sup\{\lambda(K) : K \subset J, K \text{ compact and convex}\}$.

ASSUMPTION 2. *For each n, the measure μ_n is regular.*

ASSUMPTION 3. *For each n, the measure μ_n is convex-regular.*

The above assumptions are stated here in forms convenient for immediate use in Sections 2 to 5. Certain more familiar or more readily verifiable conditions are considered in an Appendix in the course of the proof of the following:

LEMMA 1.1. *Suppose that there exists a closed convex set $V_1 \subset V$ such that V_1 is a Polish space in its relative topology and such that $X_n(\omega) \in V_1$ for all n and ω. Then Assumptions* 1, 2, *and* 3 *hold. Moreover, $s(J)$ exists and equals* $\operatorname{lan}(J)$ *for every convex J which is a relatively open subset of V_1.*

Lemma 1.1 (supplemented on occasion by Theorem 5.1) is adequate for all the examples considered in this paper.

2. Existence and properties of Lanford's entropy functions.

We begin with the following simple but useful

LEMMA 2.1. (a) *If $A \subset B$, and $s(A), s(B)$ exist, then $s(A) \leq s(B)$.* (b) *If $s(A_i)$ exists for $i = 1, \cdots, k$ then $s(\cup_{i=1}^k A_i)$ exists and equals* $\max\{s(A_i): 1 \leq i \leq k\}$.

PROOF. (a) $A \subset B$ implies $a_n(A) \leq a_n(B)$ for all n; hence $s(A) \leq s(B)$. (b) Let $B = \cup_{i=1}^k A_i$. Then $A_j \subset B$ and $\mu_n(A_j) \leq \mu_n(B) \leq k \max\{\mu_n(A_i): 1 \leq i \leq k\}$ for each j and n; hence $a_n(A_j) \leq a_n(B) \leq \max\{a_n(A_i): 1 \leq i \leq k\} + (\log k)/n$ for each j and n. Hence $s(A_j) \leq \liminf_{n\to\infty} a_n(B)$ for each $j = 1, \cdots, k$, and $\limsup_{n\to\infty} a_n(B) \leq \max\{s(A_j): 1 \leq j \leq k\}$. □

Let J be a Borel measurable convex set. Let m and n be positive integers, and let $Y = \sum_{i=m+1}^{m+n} X_i/n$. Since $\bar{X}_{m+n} = \alpha\bar{X}_m + (1-\alpha)Y$ with $\alpha = m/(m+n)$, it follows from the convexity of J that the event $\{\bar{X}_m \in J \text{ and } Y \in J\}$ implies $\{\bar{X}_{m+n} \in J\}$. Since $\bar{X}_m$ and Y are independent, and since Y has the same distribution as $\bar{X}_n$, it follows that

$$\mu_{m+n}(J) \geq \mu_m(J) \cdot \mu_n(J) \tag{2.1}$$

for all $m, n = 1, 2, \cdots$.

LEMMA 2.2. *If $J \subset V$ is a Borel measurable convex set such that $\mu_n(J) > 0$ for all $n \geq k$ then $s(J)$ exists and*

$$s(J) = \sup\{a_n(J) : n \geq k\}. \tag{2.2}$$

PROOF. It follows from (2.1) that $\log \mu_n(J)$ is finite and superadditive for $n \geqslant k$, and the lemma follows as in Lanford (1973), page 16, Lemma A2.4. □

Let S be the support of the measure μ defined by (1.5), i.e., $v \in S$ if and only if every open neighborhood of v has positive μ-measure. Let $\tilde{S}$ be the convex hull of S, and let C be the closure of $\tilde{S}$; C is called the closed convex hull of S.

LEMMA 2.3. *The set S is closed, $\mu(S) = \mu(C) = 1$, and every closed set of μ-measure 1 contains S.*

PROOF. Suppose $x \in S'$. There exists an open neighborhood of x, say A, such that $\mu(A) = 0$. By the definition of S, $v \in A$ implies $v \notin S$, so $A \subset S'$. Since x is arbitrary, S' is open, and so S is closed. Now let $K \subset S'$ be compact. Then $v \in K$ implies that there exists an open set containing v of μ-measure zero. It follows that there exist a finite number of open sets, each of μ-measure zero, whose union contains K. Hence $\mu(K) = 0$. The assumed regularity of μ now implies $\mu(S') = 0$; that $\mu(C) = 1$ follows trivially from $S \subset C$. Suppose D is closed and $\mu(D) = 1$. Then $x \in D'$ implies that D' is an open neighborhood of x of μ-measure 0, so $x \in S'$; hence $D' \subset S'$. □

For n a positive integer, let S_n be the support of the measure μ_n defined by (1.2), let $\tilde{S}_n$ be the convex hull of S_n and let C_n be the closure of $\tilde{S}_n$. The following lemma is a version of results of Barndorff-Nielsen (1978), page 91:

LEMMA 2.4. (a) *The set S_n is closed, and $\mu_n(S_n) = \mu_n(C_n) = 1$.* (b) *$S_n \supset S$ and $C_n = C$.* (c) *If J is an open set then $J \cap C \neq \varnothing$ if and only if $J \cap S_n$ is nonempty for all sufficiently large n.*

PROOF. Part (a) follows from the regularity of μ_n by replacing μ, S, and C with μ_n, S_n, and C_n in the proof of Lemma 2.3. Now let x be a point in S, and let J be an open set containing x. There exists an open convex set $J_1 \subset J$ which contains x. Then $\mu(J_1) > 0$. Hence, by an application of (2.1), $\mu_n(J_1) \geqslant [\mu(J_1)]^n > 0$; hence $\mu_n(J) > 0$. Since J is arbitrary, $x \in S_n$. Thus $S \subset S_n$; hence $\tilde{S} \subset \tilde{S}_n$ and $C \subset C_n$. We observe next that (2.1) with $J = C$ implies that $\mu_n(C) \geqslant [\mu(C)]^n = 1$; hence (by the last part of Lemma 2.3 with μ replaced with μ_n) $S_n \subset C$. Since C is convex and closed, it follows that $C_n \subset C$ and part (b) is established.

To prove (c), choose and fix an open J. If $J \cap S_n \neq \varnothing$ for some n then $J \cap C_n \neq \varnothing$ for that n; hence $J \cap C \neq \varnothing$ by part (b). Suppose now that $J \cap C \neq \varnothing$. Then $J \cap \tilde{S}$ is nonempty. Let x be a point in $J \cap \tilde{S}$. There exist positive constants $\alpha_1, \cdots, \alpha_k$ and vectors $y_1, \cdots, y_k$ in S such that $x = \sum_{i=1}^k \alpha_i y_i$, $\sum_{i=1}^k \alpha_i = 1$. For each $n = 1, 2, \cdots$ let $j_{in} = [n\alpha_i]$ and $\alpha_{in} = j_{in}/n$ for $1 \leqslant i \leqslant k-1$, and let $\alpha_{kn} = 1 - \sum_{i=1}^{k-1} \alpha_{in}$, $j_{kn} = n\alpha_{kn}$. Then $\alpha_{in} \to \alpha_i$ as $n \to \infty$ for $1 \leqslant i \leqslant k$; hence $x_n = \sum_{i=1}^k \alpha_{in} y_i \to x$. Hence, for all sufficiently large n, $x_n \in J$ and $j_{in} \geqslant 1$ for $1 \leqslant i \leqslant k$. Choose and fix such an n. For each i let B_{in} be an open *convex* neighborhood of y_i such that

$$\alpha_{1n} B_{1n} + \cdots + \alpha_{kn} B_{kn} \subset J. \tag{2.3}$$

Since $y_i \in S$, $\mu(B_{in}) > 0$ for $1 \leqslant i \leqslant k$. Let E_n denote the event that in the sequence $X_1, \cdots, X_n$ the first j_{1n} random vectors lie in B_{1n}, the next j_{2n} lie in $B_{2n}, \cdots$, and the last j_{kn} lie in B_{kn}. Then, by (1.16), $P(E_n) > 0$, and E_n implies that $\bar{X}_n \in J$, by (2.3). Hence $P(\bar{X}_n \in J) = \mu_n(J) > 0$; hence $\mu_n(J \cap S_n) > 0$ by part (a); hence $J \cap S_n \neq \varnothing$. □

THEOREM 2.1. *If J is a finite union of open convex sets then $s(J)$ exists, $-\infty \leqslant s(J) \leqslant 0$, and $s(J) > -\infty$ if and only if $J \cap C \neq \varnothing$.*

PROOF. In view of Lemma 2.1(b), it will suffice to consider a single open convex J. Suppose first that $J \cap C = \varnothing$. Then $\mu_n(J) = 0$ by Lemma 2.4(a), (b); hence $a_n(J) = -\infty$ for every n, and $s(J) = -\infty$. Suppose now that $J \cap C \neq \varnothing$. It then follows from Lemma 2.4(c) that there exists a positive integer k such that $\mu_n(J) > 0$ for all $n \geqslant k$. The existence and finiteness of $s(J)$ now follows from Lemma 2.2. □

For each $v \in V$, let $s(v)$ be defined by (1.6).

THEOREM 2.2. *The function s: $V \to [-\infty, 0]$ is concave and upper semicontinuous.*

PROOF. To prove upper semicontinuity, choose $x \in V$, and let r be a real number such that $s(x) < r$. Then, whether $s(x) > -\infty$ or $= -\infty$, there exists an open convex set J containing x such that $s(J) < r$; hence $s(v) < r$ for every $v \in J$, by (1.6). Thus s is upper semicontinuous at x.

Now choose $x, y \in V$, and α, $0 < \alpha < 1$, and let $z = \alpha x + (1 - \alpha)y$. Let J be an open convex set containing z. Then there exist open convex sets containing x and y respectively, say J_1 and J_2, such that $\alpha J_1 + (1 - \alpha)J_2 \subset J$.

Suppose first that α is rational, say $\alpha = j/k$, $1 \leqslant j < k$. For r a positive integer, let $Y = \sum_{i=jr+1}^{kr} X_i/(k - j)r$. It follows from the preceding paragraph that the event $\{\bar{X}_{jr} \in J_1$ and $Y \in J_2\}$ implies $\{\bar{X}_{kr} \in J\}$. Since Y and $\bar{X}_{jr}$ are independent, and Y has the same distribution as $\bar{X}_{(k-j)r}$, it follows that $\mu_{jr}(J_1) \cdot \mu_{(k-j)r}(J_2) \leqslant \mu_{kr}(J)$. Hence $\alpha a_{jr}(J_1) + (1 - \alpha)a_{(k-j)r}(J_2) \leqslant a_{kr}(J)$. Letting $r \to \infty$, we obtain $s(J) \geqslant \alpha s(J_1) + (1 - \alpha)s(J_2)$. Hence $s(J) \geqslant \alpha s(x) + (1 - \alpha)s(y)$ by (1.6). Since J is arbitrary, it follows that

$$s(\alpha x + (1 - \alpha)y) \geqslant \alpha s(x) + (1 - \alpha)s(y). \tag{2.4}$$

Thus (2.4) holds for every rational α; that it holds for all $\alpha \in (0, 1)$ now follows from the upper semicontinuity of s. □

Our next objective is to show that (1.8) holds for open convex sets. As noted previously, if J is open (1.12) implies that $\liminf_{n\to\infty} a_n(J) \geqslant \operatorname{lan}(J)$; the problem is to show that (1.13) holds. Lanford's method is, in effect, to show that (1.13) does hold for compact sets J, and then to show that it holds for an open convex set by approximating the set by compact convex sets.

LEMMA 2.5. *If $K \subset V$ is a compact set then*

$$\limsup_{n\to\infty} a_n(K) \leqslant \operatorname{lan}(K). \tag{2.5}$$

PROOF. Since (2.5) holds trivially if $K = \varnothing$, assume that K is nonempty. Let $\varepsilon > 0$ and $m > 0$ be constants. It follows from the definition of the point entropy

$s(v)$ that for each $v \in K$ there exists an open convex set containing v, A say, such that, whether $s(v) = -\infty$ or $> -\infty$, $s(A) \leq \max\{-m, s(v) + \varepsilon\} \leq \max\{-m, \operatorname{lan}(K) + \varepsilon\}$. Since K is compact there exist open convex sets $A_1, \cdots, A_k$ such that $K \subset \cup_1^k A_i = B$ say, and $s(A_i) \leq \max\{-m, \operatorname{lan}(K) + \varepsilon\}$ for $i = 1, \cdots, k$. Since $a_n(K) \leq a_n(B)$ for all n, and since each A_i has an entropy, it follows from Lemma 2.1(b) that $\lim\sup_{n\to\infty} a_n(K) \leq s(B) = \max\{s(A_i): 1 \leq i \leq k\}$; hence $\lim\sup_{n\to\infty} a_n(K) \leq \max\{-m, \operatorname{lan}(K) + \varepsilon\}$. If $\operatorname{lan}(K) = -\infty$, (2.5) follows by letting $m \to \infty$; if $\operatorname{lan}(K) > -\infty$, (2.5) follows by letting $\varepsilon \to 0$ and then choosing m sufficiently large. □

LEMMA 2.6. *If $K \subset V$ is a closed convex set then* (2.5) *holds.*

This lemma is not used in the present paper and its proof is omitted. We note here in connection with Lemmas 2.5, 2.6, and 2.7 that in general $s(K)$ does not exist for every compact convex K (cf. Example 6.1).

LEMMA 2.7. *If J is an open convex set with $s(J) > -\infty$ then for given $\varepsilon > 0$ there exists a compact convex $K \subset J$ such that $s(K)$ exists and $s(K) \geq s(J) - \varepsilon$.*

PROOF. Since $s(J) > -\infty$ there exists a positive integer k such that $\mu_n(J) > 0$ for all $n \geq k$; we may and do assume that $a_k(J) \geq s(J) - (\varepsilon/2)$. Since each μ_n is convex-regular, for each integer n with $k \leq n \leq 2k$ there exists a compact convex $L_n \subset J$ such that $a_n(L_n) \geq a_n(J) - (\varepsilon/2) > -\infty$. Let $K_1 = \cup_{n=k}^{2k} L_n$. Then K_1 is compact, $K_1 \subset J$, $a_k(K_1) \geq s(J) - \varepsilon$, and $a_n(K_1) > -\infty$ for $k \leq n \leq 2k$. Let K be the convex hull of K_1; it follows from Choquet ((1969), I, page 337) by the convexity of J that K has all the properties just listed for K_1; moreover, since K is convex, it follows from (2.1) (with $J = K$) that $\mu_n(K) > 0$ for all $n \geq k$. It now follows from Lemma 2.2 (with $J = K$) that $s(K)$ exists and $s(K) \geq a_k(K) \geq s(J) - \varepsilon$. □

THEOREM 2.3. *If J is a finite union of open convex sets then* $s(J) = \operatorname{lan}(J)$.

PROOF. It will suffice to show that if J is an open convex set then $s(J) \leq \operatorname{lan}(J)$. Since this holds trivially if $s(J) = -\infty$, assume $s(J) > -\infty$. Choose $\varepsilon > 0$. It follows from Lemma 2.7 that there exists a compact $K \subset J$ such that $s(K) \geq s(J) - \varepsilon$; since $\operatorname{lan}(K) \leq \operatorname{lan}(J)$, Lemma 2.5 implies $s(J) \leq \operatorname{lan}(J) + \varepsilon$. □

COROLLARY 2.1. *Either $s(v) = 0$ for all v in V, or*

$$\sup\{s(v) : v \in V\} = 0 \quad \text{and} \quad \inf\{s(v) : v \in V\} = -\infty. \tag{2.6}$$

PROOF. V is an open convex set with $\mu_n(V) = 1$ for all n; hence $s(V) = 0$. It follows from Theorem 2.3 that $\operatorname{lan}(V) = 0$, i.e., the first part of (2.6) holds. Suppose $s \not\equiv 0$. Then there exists an x_1 with $s(x_1) < 0$. If $s(x_1) = -\infty$ the corollary is established. Suppose then that $s(x_1) > -\infty$. The first part of (2.6) implies the existence of x_2 such that $-\infty < s(x_1) < s(x_2) \leq 0$. Let α be a constant, $0 < \alpha < 1$, and let x_3 be determined by $\alpha x_2 + (1 - \alpha)x_3 = x_1$. The concavity of s implies $\alpha s(x_2) + (1 - \alpha)s(x_3) \leq s(x_1)$, and it is plain that $s(x_3) \to -\infty$ as $\alpha \to 1$. □

Let F be the set of all $v \in V$ such that $s(v) > -\infty$. Let F° be the (possibly empty) interior of F, and let I be the (possibly empty) interior of C.

THEOREM 2.4. (a) $F \subset C$. (b) *If* F° *is nonempty then*

$$F^\circ = I \quad \text{and} \quad \bar{F} = C. \tag{2.7}$$

(c) *If* V *is finite-dimensional and* I *is nonempty then* (2.7) *holds and* s *is continuous on* I.

PROOF. Suppose $v \in C'$. There exists an open convex neighborhood of v, J say, with $J \subset C'$. Then $s(J) = -\infty$ by Theorem 2.1; hence $s(v) = -\infty$ by (1.6). Thus $C' \subset F'$, so part (a) holds. To establish part (b) we note first that, since s is concave, $F = \{w : s(w) > -\infty\}$ is a convex set. Let $v \in V$ be such that $s(v) = -\infty$. Then $v \notin F$, and since F° is nonempty there exists (Choquet (1969), II, page 30) a nonzero $\theta \in V^*$ and a constant γ such that $\theta(v) \leqslant \gamma$ and $\theta(w) \geqslant \gamma$ for every $w \in F$. Let $L = \{w : \theta(w) < \gamma\}$. Then $s(w) = -\infty$ for all $w \in L$; hence $s(L) = -\infty$ by Theorem 2.3; hence $L \subset C'$ by Theorem 2.1. Since $v \in \bar{L}$, either $v \notin C$ or v is on the boundary of C; in either case, $v \notin I$. Thus $\{v : s(v) = -\infty\} \subset I'$; hence $I \subset F$; hence $I \subset F^\circ$. Part (a) implies $F^\circ \subset C$ and so $F^\circ \subset I$. Thus $F^\circ = I$, and so I is nonempty. Hence $\bar{I} = C$ (Choquet (1969), I, page 335); the second part of (2.7) now follows from

$$I \subset F \subset C \tag{2.8}$$

and part (b) is established. Now choose $v \in V$, and for $\theta \in V^*$ let $H_\theta(v) = \{w : \theta(w) \geqslant \theta(v)\}$. As is noted in the first paragraph of Section 3, $s(H_\theta)$ exists. It follows from (2.1) that $\mu_n(H_\theta) \geqslant [\mu(H_\theta)]^n$ for every n; hence

$$s(v) \geqslant \inf\{\log \mu(H_\theta(v)) : \theta \in V^*\} \tag{2.9}$$

by Theorems 3.1 and 3.2. To prove part (c) suppose that V is finite-dimensional and I is nonempty. It then follows from (2.9) by Lemma 9.2 of Barndorff-Nielsen (1978) that $s(v) > -\infty$ for $v \in I$, so $I \subset F$. Since I is open, F° is nonempty, and it follows from part (b) that (2.7) holds. That s is continuous on I follows from Choquet (1969), I, page 343.

In Theorem 2.4 the hypothesis that F° be nonempty is essential to part (b), and the hypothesis that V be finite-dimensional is essential to part (c); this may be seen from Examples 6.3 and 6.4. The following is a finite-dimensional example in which both inclusions in (2.8) are strict and s is not continuous on F.

EXAMPLE 2.1. Suppose $V = R^2$ is the Euclidean plane. Let D be the closed unit disc $\{v : \|v\| \leqslant 1\}$, and let λ denote the uniform probability measure on D, i.e., $\lambda(A) = (\text{constant}) \times$ Lebesgue measure of $A \cap D$. Let $v_0 = (0, 1)$, let ν be the probability measure degenerate at v_0, and let $\mu = (\lambda + \nu)/2$. Then $S = \tilde{S} = C = D$, and $I = \{v : \|v\| < 1\}$. It follows easily from Theorems 2.4, 3.1, and 3.2 that $s(v) > -\infty$ for $v \in I$; that $\|v\| \geqslant 1$ implies $s(v) = -\infty$ unless $v = v_0$; and that $s(v_0) = \log(\frac{1}{2})$. Hence $F = I \cup \{v_0\}$. To see that s is not continuous on F choose t, $-\infty < t < \log(\frac{1}{2})$. For each $k = 1, 2, \cdots$ let v_k be such that $\|v_k\| = 1$ and $\|v_k - v_0\| = k^{-1}$. Since $s(v_k) = -\infty$, it follows from the definition of s that there exists $r_k > 0$ such that $\|v - v_k\| < r_k$ implies $s(v) < t$. Let w_k be a point in I such that $\|w_k - v_k\| < \min\{k^{-1}, r_k\}$. Then $\{w_k : k = 1, 2, \cdots\}$ is a sequence in I such that $s(w_k) < t$ and $\|w_k - v_0\| < 2k^{-1}$ for each k. Hence $w_k \to v_0$ as $k \to \infty$, but $\lim\sup_{k\to\infty} s(w_k) < s(v_0)$.

3. Some formulae concerning the point entropy. For $v \in V, \theta \in V^*$, and for $\varepsilon > 0$ let

$$(3.1) \qquad L_\theta(v, \varepsilon) = \{w : \theta(w) > \theta(v) - \varepsilon\}.$$

It follows from Theorems 2.1, 2.3 that $s(L_\theta)$ exists and equals $\operatorname{lan}(L_\theta)$. We observe next that

$$(3.2) \qquad H_\theta(v) = \{w : \theta(w) \geqslant \theta(v)\}$$

also has an entropy. To see this, let $Y_n = \theta(X_n)$ for each n. Then $\bar{X}_n \in H_\theta(v)$ if and only if $\bar{Y}_n \geqslant \theta(v)$. It now follows from Chernoff's theorem that $s(H_\theta)$ exists and (cf. (1.4), (1.9)) that

$$(3.3) \qquad s(H_\theta(v)) = \inf\{-t\theta(v) + c(t\theta) : t \geqslant 0\}$$

where $(t\theta)(w) = t \cdot \theta(w)$. A relatively simple proof of Chernoff's theorem is given in Bahadur (1971). It can be shown that we also have $s(H_\theta) = \operatorname{lan}(H_\theta)$.

If θ is not the zero functional, $H_\theta(v)$ is the typical closed half-space with v on its boundary, and $L_\theta(v, \varepsilon)$ is the typical open half-space containing v. The following theorem in effect relates the Fenchel transform of the cumulant generating function of μ to the entropies of half-spaces.

THEOREM 3.1. *For each v in V,*

$$(3.4) \qquad -c^*(v) = \inf\{s(H_\theta(v)) : \theta \in V^*\}$$

and

$$(3.5) \qquad -c^*(v) = \inf\{s(L_\theta(v, \varepsilon)) : \theta \in V^*, 0 < \varepsilon < \infty\}.$$

PROOF. The range of $t\theta$ as t varies over $[0, \infty)$ and θ over V^* is V^*. It follows hence from (1.10) and (3.3) that (3.4) holds. To establish (3.5), choose $\theta \in V^*, \varepsilon > 0$, and for each n let $Y_n = \theta(X_n) - \theta(v) + \varepsilon$, and let $T_n = \sum_{i=1}^n Y_i$. Then $\bar{X}_n \in L_\theta(v, \varepsilon)$ if and only if $T_n > 0$. Hence $\mu_n(L_\theta(v, \varepsilon)) = P(T_n > 0) \leqslant P(\exp(T_n) \geqslant 1) \leqslant E(\exp T_n) = \exp\{n[c(\theta) - \theta(v) + \varepsilon]\}$; hence $a_n(L_\theta(v, \varepsilon)) \leqslant c(\theta) - \theta(v) + \varepsilon$; hence $s(L_\theta(v, \varepsilon)) \leqslant c(\theta) - \theta(v) + \varepsilon$; hence, with $r(v)$ the right-hand side of (3.5), $r(v) \leqslant c(\theta) - \theta(v) + \varepsilon$. Since θ and ε are arbitrary, $r(v) \leqslant -c^*(v)$ by (1.10). Since $L_\theta(v, \varepsilon) \supset H_\theta(v)$ for every θ and $\varepsilon > 0$, it follows from Lemma 2.1(a) and (3.4) that $r(v) \geqslant -c^*(v)$. □

THEOREM 3.2. *For each v in V, $s(v) = -c^*(v)$.*

PROOF. Choose and fix a $v \in V$. Let $L_\theta(v, \varepsilon)$ be defined by (3.1). It follows from the definition of $s(v)$ that $s(v) \leqslant s(L_\theta(v, \varepsilon))$. Since θ and ε are arbitrary, it follows from (3.5) that $s(v) \leqslant -c^*(v)$. It remains to show that $-c^*(v) \leqslant s(v)$.

Choose positive constants m and δ, and let $t(v) = \max\{-m, s(v) + \delta\}$. We shall show that there exists a $\theta \in V^*$ and an $\varepsilon > 0$ such that

$$(3.6) \qquad s(w) < t(v) \quad \text{for} \quad w \in L_\theta(v, \varepsilon).$$

Let $A = \{w : s(w) \geqslant t(v)\}$. If A is empty, (3.6) holds for every θ and ε; suppose then that A is nonempty. It follows from Theorem 2.2 that A is a closed convex set; moreover, whether $s(v) = -\infty$ or $> -\infty$, $v \notin A$. Consequently (Choquet (1969), II, page 28) there exist $\theta \in V^*$ and $\gamma \in R^1$ such that $\theta(v) > \gamma$ and $\theta(w) < \gamma$ for $w \in A$. Let $\varepsilon = \theta(v) - \gamma$; then (3.6) holds. It follows from (3.6) by Theorem 2.3 that $s(L_\theta(v, \varepsilon)) \leqslant t(v)$; hence, by Theorem 3.1, $-c^*(v) \leqslant t(v)$. Since m and δ in the definition of $t(v)$ are arbitrary it follows that $-c^*(v) \leqslant s(v)$. □

Theorems 3.1 and 3.2 provide three related but distinct methods of computing the point entropy: use (1.10), or (3.4), or (3.5) to find c^* and then $s = -c^*$. Next, we describe a very different method, based on the Kullback-Leibler information numbers, of finding the point entropy; this method is implicit in Sanov (1957) and is considered explicitly by Lanford (1973); cf. also Donsker and Varadhan (1976). Let ν be a probability measure on the Borel sets of V. If ν is not absolutely continuous with respect to μ let $K(\nu) = +\infty$; if ν is absolutely continuous, let $K(\nu) = \int_V \log[(d\nu/d\mu)(v)]\nu(dv)$, $0 \leqslant K \leqslant \infty$. In the following if v is a point in V and ν is a probability measure, we write $E(X_1|\nu) = v$ if and only if for every $\theta \in V^*$ the integral $\int_V \theta(w)\nu(dw)$ exists and equals $\theta(v)$.

Let $M(v)$ be the set of all probability measures ν such that $E(X_1|\nu) = v$. $M(v)$ is nonempty since $\{v\}$ is Borel measurable and the measure which assigns mass 1 to $\{v\}$ is in $M(v)$. Let

$$\sigma(v) = \sup\{-K(\nu) : \nu \in M(v)\}. \tag{3.7}$$

The following Theorem 3.3 and Example 3.2 at the end of this section show that although σ is not quite the point entropy function it comes very close; in particular σ can be used instead of s in computing the entropy of any open convex set. According to part (c) of Theorem 3.3 the technical reason why σ fails to coincide with s is because in general σ is not upper semicontinuous.

THEOREM 3.3.
(a) *The function* $\sigma : V \to [-\infty, 0]$ *is concave.*
(b) *For all* v *in* V, $\sigma(v) \leqslant s(v)$.
(c) *For all* v *in* V, $\limsup_{w \to v} \sigma(w) = s(v)$.
(d) *For every nonempty open set* J, $\sup\{\sigma(v) : v \in J\} = \text{lan}(J)$.
(e) *If* V *is finite-dimensional,* $\{v : \sigma(v) \neq s(v)\}$ *is a subset of the boundary of* C.
(f) *If* V *is one-dimensional,* $\sigma(v) = s(v)$ *for all* v *in* V.

At present we shall establish only part (b). Choose $\theta \in V^*$, $v \in V$, and let $H_\theta(v)$ be defined by (3.2). It follows from Theorem 4.2 in Bahadur (1971) that $s(H_\theta(v))$ equals the supremum of $-K(\nu)$ over all ν such that $i = \int_V \theta(w)\nu(dw)$ exists and $\theta(v) \leqslant i \leqslant \infty$. This last set of measures contains $M(v)$; hence $\sigma(v) \leqslant s(H_\theta(v))$ by (3.7). Since θ is arbitrary, it follows from Theorems 3.1 and 3.2 that $\sigma(v) \leqslant s(v)$, so part (b) holds. The rest of the proof of Theorem 3.3 is deferred to Section 4. An interesting structural relation between the functions s and σ is pointed out in Section 5.

COROLLARY 3.1. *If* $E(X_1|\mu) = v$ *then* $s(v) = 0$.

PROOF. The hypothesis $\mu \in M(v)$ inplies $\sigma(v) \geqslant -K(\mu) = 0$; hence $\sigma(v) = s(v) = 0$ by Theorem 3.3(b). ∎

Corollary 3.1 is immediate in case the law of large numbers holds, i.e., $E(X_1 | \mu)$ exists and equals v, say, and $\lim_{n\to\infty} P(\bar{X}_n \in J) = 1$ for every open J containing v; for then $s(J) = 0$ for every such J, and hence $s(v) = 0$ by (1.6).

For each θ in the natural parameter space Θ let μ_θ be the measure on V defined by $\mu_\theta(dw) = \exp[\theta(w) - c(\theta)]\mu(dw)$. For each n and θ let $\mu_{n,\theta}$ be the probability measure on V defined by

$$\mu_{n,\theta}(dw) = \exp[n\theta(w) - nc(\theta)]\mu_n(dw). \tag{3.8}$$

The measure $\mu_{n,\theta}$ is to be thought of as the distribution of $\bar{X}_n$ when the marginal distribution of the independent X_i is μ_θ. Since $\mu_{n,\theta}$ is dominated by μ_n, $\mu_{n,\theta}$ is also regular and convex-regular. It can be shown that in fact the sequence $\{\mu_{n,\theta}\}$ has all the properties of $\{\mu_n\}$; in particular, $\lim_{n\to\infty} n^{-1}\log \mu_{n,\theta}(J)$ exists for every open convex J, say $s_\theta(J)$. Let $s_\theta(v)$ be defined by (1.6) with $s(J)$ replaced by $s_\theta(J)$. We note that our μ, $s(J)$, and $s(v)$ are μ_0, $s_0(J)$, and $s_0(v)$.

THEOREM 3.4. *For every v in V and θ in Θ*

$$s_\theta(v) = s(v) + \theta(v) - c(\theta). \tag{3.9}$$

PROOF. Choose $v \in V, \theta \in \Theta$, and $\varepsilon > 0$. Let $J_1 = \{w : |\theta(w) - \theta(v)| < \varepsilon\}$. Let J be an open convex neighborhood of v. It follows easily from (3.8) that, for every n,

$$\mu_{n,\theta}(J \cap J_1) = \exp\{n\theta(v) - nc(\theta) + n\delta_n\varepsilon\} \cdot \mu_n(J \cap J_1) \tag{3.10}$$

where $|\delta_n| \leqslant 1$. Since $s(J \cap J_1)$ and $s_\theta(J \cap J_1)$ both exist, it follows from (3.10) that

$$\begin{aligned} s_\theta(J \cap J_1) &\leqslant s(J \cap J_1) + \theta(v) - c(\theta) + \varepsilon \\ s_\theta(J \cap J_1) &\geqslant s(J \cap J_1) + \theta(v) - c(\theta) - \varepsilon. \end{aligned} \tag{3.11}$$

It is plain that $s(v)[s_\theta(v)]$ is the infimum of $s(J \cap J_1)[s_\theta(J \cap J_1)]$ over all open convex sets J containing v. It follows hence from (3.11) first that $s_\theta(v) \leqslant s(J \cap J_1) + \theta(v) - c(\theta) + \varepsilon$ and $s_\theta(J \cap J_1) \geqslant s(v) + \theta(v) - c(\theta) - \varepsilon$, and next that $s_\theta(v) \leqslant s(v) + \theta(v) - c(\theta) + \varepsilon$ and $s_\theta(v) \geqslant s(v) + \theta(v) - c(\theta) - \varepsilon$. Since ε is arbitrary it follows that (3.9) holds. ∎

COROLLARY 3.2. *For each v in V*

$$\sup\{s_\theta(v) : \theta \in \Theta\} = 0 \ \text{or} \ -\infty. \tag{3.12}$$

PROOF. If $s(v) = -\infty$ it follows from (3.9) that $s_\theta(v) = -\infty$ for all $\theta \in \Theta$. If $s(v) > -\infty$, (3.9) implies that the supremum in (3.12) equals $s(v) + c^*(v)$, and this is zero by Theorem 3.2. ∎

It follows from (3.8) that $\log(d\mu_\theta/d\mu)(w) = \theta(w) - c(\theta)$; hence, for $\theta \in \Theta$,

$$\begin{aligned} K(\mu_\theta) &= \int_V \theta(w)\mu_\theta(dw) - c(\theta) \\ &= \int_V \theta(w)\exp[\theta(w) - c(\theta)]\mu(dw) - c(\theta). \end{aligned} \tag{3.13}$$

The integrals in (3.13) always exist since $K(\nu)$ is well defined for every ν but they could equal $+\infty$.

Let W be the (possibly empty) set of all v in V such that

$$E(X_1 | \mu_\theta) = v \tag{3.14}$$

for some θ in Θ. We shall call W the space of means. For each $v \in W$ let $\Theta(v)$ denote the set of all $\theta \in \Theta$ such that (3.14) holds, i.e., $\Theta(v) = \{\theta : \theta \in \Theta, \mu_\theta \in M(v)\}$.

COROLLARY 3.3. *Suppose that v is a point in W. Then*

$$s(v) = \sigma(v) > -\infty \tag{3.15}$$

and

$$s(v) = -K(\mu_\theta) = -\theta(v) + c(\theta) \tag{3.16}$$

for every θ in $\Theta(v)$. If θ and δ are in $\Theta(v)$ then

$$\mu_\theta = \mu_\delta \tag{3.17}$$

on the Borel sets of V.

PROOF. Choose $\theta \in \Theta(v)$. It follows from Corollary 3.1 with μ replaced by μ_θ that $s_\theta(v) = 0$; hence $s(v) = -\theta(v) + c(\theta)$ by (3.9); hence $s(v) = -K(\mu_\theta)$ by (3.13) and (3.14), and (3.16) is established. Since $\mu_\theta \in M(v)$, $\sigma(v) \geqslant -K(\mu_\theta) = s(v)$; hence (3.15) holds, by Theorem 3.3(b). Now suppose $\delta \in \Theta(v)$. Then (3.16) holds with θ replaced by δ; hence $-s(v) = K(\mu_\theta) = K(\mu_\delta) < \infty$. Suppose $\mu_\theta \neq \mu_\delta$. It is easily seen that K is a strictly convex functional on the convex set $\{\nu : K(\nu) < \infty\}$. Consequently, $\lambda = (\mu_\theta + \mu_\delta)/2$ is a measure in $M(v)$ with $K(\lambda) < -s(v)$; hence $\sigma(v) \geqslant -K(\lambda) > s(v)$, contrary to (3.15). □

It follows from the last part of Corollary 3.3 that if

$$\mu\{v : \theta(v) = \gamma\} \neq 1 \quad \text{for nonzero} \quad \theta \in V^* \quad \text{and} \quad \gamma \in R^1 \tag{3.18}$$

then, for each $v \in W$, the set $\Theta(v)$ consists of a single linear functional.

It may be worthwhile to note here the following connections with classical nonasymptotic theories of estimation and testing. Suppose that V is finite dimensional, say $V = R^k$. Then $V^* = R^k$ and Θ is a (not necessarily open) convex subset of R^k. Suppose that the unknown measure on V is some member of $\{\mu_\theta : \theta \in \Theta\}$ and suppose that the observed sample point is v. Then $c^*(v) = -s(v)$ is equivalent to the likelihood ratio statistic of Neyman and Pearson for testing $\theta = 0$, large values of c^* being significant, and the conditions (3.14) are equivalent to the likelihood equations. Corollary 3.3 implies that solutions of the likelihood equations actually maximize the likelihood function. A detailed study of maximum likelihood estimation in exponential families is given in Barndorff-Nielsen (1978); in particular, sufficient conditions for $I \subset W$ or even $W = V$ are discussed there.

The following Examples 3.1 are simple illustrations of various cases that arise in the contexts of Corollaries 2.1, 3.2, and 3.3.

EXAMPLES 3.1. In each of the examples (a)–(d), $V = R^1$, $V^* = R^1$, and $\theta(v) = \theta \cdot v$.

(a) μ assigns probability one to $\{0\}$. Then $s(0) = 0$ and $s(v) = -\infty$ for $v \neq 0$. Here $\Theta = V^*$, $W = \{0\}$, $\Theta(0) = \Theta$, and $\mu_\theta = \mu$ for all θ.

(b) $\mu(dv) = a_1/(1 + |v|^p)\, dv$ for $-\infty < v < \infty$, where $1 < p \leq 2$ and a_1 is a normalizing constant. Here $c(\theta) = +\infty$ for $\theta \neq 0$ and $c(0) = 0$, so $s(v) = -c^*(v) = 0$ for all v; $\Theta = \{0\}$, and W is empty.

(c) $\mu(dv) = a_2[1 + v^4]^{-1} \exp(-v)\, dv$ for $0 < v < \infty$, and $\mu((-\infty, 0]) = 0$. Here $\Theta = (-\infty, +1]$; $W = (0, E(X_1 | \mu_1)]$; $s = -\infty$ on $(-\infty, 0]$; s is strictly concave on $(0, E(X_1 | \mu_1)]$; s increases from $-\infty$ to zero over $(0, E(X_1 | \mu)]$ and then decreases to a negative value at $E(X_1 | \mu_1)$; and s decreases linearly to $-\infty$ over $[E(X_1 | \mu_1), +\infty)$.

(d) $\mu(dv) = (2\pi)^{-\frac{1}{2}} \exp(-v^2/2)\, dv$ for $-\infty < v < \infty$. Here $s(v) = -v^2/2$ for each v, $\Theta = V^*$, and $W = V$.

The following example shows that (3.15) can break down even in two dimensions (cf. Theorem 3.3). The example is based on a counterexample constructed by Barndorff-Nielsen (1978), page 174, for a different purpose.

EXAMPLE 3.2. Suppose $V = R^2$. Let $p_0, p_1, \cdots$ be positive numbers with $\Sigma_{j=0}^{\infty} p_j = 1$. For each $j = 2, 3, \cdots$ choose $a_j > 1$ so that $a_{j+1} > a_j$ for all j and

$$\Sigma_{j=2}^{\infty} p_j \exp(x a_j) = +\infty \quad \text{for all} \quad x > 0, \tag{3.19}$$

e.g., $a_2 = 2$ and $a_{j+1} = \max\{a_j + 1, 1/p_{j+1}\}$ for all $j \geq 2$. Let $v_0 = (0, 0)$, $v_1 = (1, 0)$, and for $j \geq 2$ let $v_j = (a_j, 1)$. Let μ be defined by $\mu(\{v_j\}) = p_j$ for $j = 0, 1, 2, \cdots$.

The only probability measure ν which is dominated by μ and satisfies $E(X_1 | \nu) = v_1$ is the measure degenerate at v_1. It follows hence that

$$\sigma(v_1) = \log p_1. \tag{3.20}$$

We observe next that the cumulant generating function of μ evaluated at $(x, y) \in V^* = R^2$ is

$$c(x, y) = \log\left[p_0 + p_1 \exp(x) + \Sigma_{j=2}^{\infty} p_j \exp(x a_j + y)\right]. \tag{3.21}$$

It follows easily from (3.19) and (3.21) that $c^*(v_1) = \sup\{x - c(x, y) : (x, y) \in R^2\}$ equals $-\log(p_0 + p_1)$. Hence

$$s(v_1) = \log(p_0 + p_1) \tag{3.22}$$

by Theorem 3.2. It is plain from (3.20) and (3.22) that $s(v_1) > \sigma(v_1) > -\infty$.

In the present notation, Theorem 5.2 of Donsker and Varadhan (1976) states that if V is a Banach space, and if (1.15) holds, then (i) σ is an upper semicontinuous function on V into $[-\infty, 0]$; (ii) $\sigma(v) = 0$ if and only if $E(X_1 | \mu) = v$; (iii) for each real t, $\{v : \sigma(v) \geq t\}$ is a compact set; and (iv) $\sigma(v) = -c^*(v)$ for all v. In view of Theorem 3.2, this theorem is a useful complement to Theorem 3.3 and Corollary 3.1. It can be seen from Examples 3.1(b) and 3.2 that the condition (1.15) is essential to every part of the stated theorem of Donsker and Varadhan.

4. Proof of Theorem 3.3 (Continued). It is well known and easy to see that $K(\nu)$ is a convex functional on the set of probability measures on V. Let x, y be points in V, let $\alpha \in (0, 1)$, and let $z = \alpha x + (1 - \alpha)y$. Let ν_1 and ν_2 be measures in $M(x)$ and $M(y)$ respectively. Then $\alpha\nu_1 + (1 - \alpha)\nu_2$ is a measure in $M(z)$, and hence $\sigma(z) \geqslant -K(\alpha\nu_1 + (1 - \alpha)\nu_2) \geqslant \alpha(-K(\nu_1)) + (1 - \alpha)(-K(\nu_2))$. Since ν_1 and ν_2 are arbitrary, $\sigma(z) \geqslant \alpha\sigma(x) + (1 - \alpha)\sigma(y)$, so part (a) holds. Part (b), i.e.,

$$\sigma(v) \leqslant s(v) \tag{4.1}$$

for all v in V, has already been established in Section 3.

Now regard V as a sample space equipped with the probability measure μ, and let $Y(v)$ be a real valued measurable function on V. Let N be the set of all ν such that $i = \int_V Y(v)\nu(dv)$ exists and $0 < i \leqslant \infty$. Let $b = \sup\{-K(\nu) : \nu \in N\}$ if $N \neq \varnothing$ and let $b = -\infty$ if $N = \varnothing$. Let $Y_1, Y_2, \cdots$ denote a sequence of independent replicates of Y, and for each $n = 1, 2, \cdots$ let $b_n = n^{-1} \log P(Y_1 + \cdots + Y_n > 0)$.

LEMMA 4.1. $\lim_{n\to\infty} b_n = b$.

PROOF. Suppose $\mu(Y \leqslant 0) = 1$. Then no measure in N is dominated by μ; hence $b_n = b = -\infty$ for each n, and the lemma holds. Suppose then that $\mu(Y > 0) > 0$. In this case $N \neq \varnothing$. Let $\beta_n(u) = n^{-1} \log P(Y_1 + \cdots + Y_n \geqslant u)$, and let $\beta(u) = \lim_{n\to\infty} \beta_n(u)$; $\beta(u)$ exists for every real u, by Chernoff's theorem. It follows from Lemma 3.3 of Bahadur (1971) that $\beta(u)$ is finite and continuous in a neighborhood of $u = 0$. Choose $\varepsilon > 0$. There exists $t > 0$ such that $\beta(t) \geqslant \beta(0) - \varepsilon$. It follows from Theorem 4.2 in Bahadur (1971) and the definition of b that

$$b \leqslant \beta(0), \quad b \geqslant \beta(t) \geqslant \beta(0) - \varepsilon. \tag{4.2}$$

We observe next that $\beta_n(t) \leqslant b_n \leqslant \beta_n(0)$ for every n. Hence

$$\liminf_{n\to\infty} b_n \geqslant \beta(t), \limsup_{n\to\infty} b_n \leqslant \beta(0). \tag{4.3}$$

Since ε is arbitrary, it follows from (4.2) and (4.3) that $b = \beta(0)$ and that Lemma 4.1 holds. □

Now choose and fix $x \in V$, $\theta \in V^*$, and $\varepsilon > 0$, and let $L_\theta(x, \varepsilon)$ be defined by (3.1). In the following, $L_\theta(x, \varepsilon)$ is often abbreviated to L_θ.

LEMMA 4.2. $s(L_\theta) = \sup\{\sigma(v) : v \in L_\theta\}$.

PROOF. Suppose first that $L_\theta \cap C = \varnothing$. Then $s(L_\theta) = -\infty$ by Theorem 2.1; hence $s = -\infty$ on L_θ by Theorem 2.3; hence $\sigma = -\infty$ on L_θ by (4.1), and Lemma 4.2 holds. Suppose then that $L_\theta \cap C \neq \varnothing$. Then $s(L_\theta) > -\infty$. Let $Y(v) \equiv \theta(v) - \theta(x) + \varepsilon$ in Lemma 4.1, and let N be the set of measures determined as above by this Y. Then $N \neq \varnothing$ and $s(L_\theta) = \sup\{-K(\nu) : \nu \in N\}$ by Lemma 4.1. Since $s(L_\theta) > -\infty$ it follows that

$$s(L_\theta) = \sup\{-K(\nu) : \nu \in N_1\} \tag{4.4}$$

where $N_1 = \{\nu : \nu \in N, K(\nu) < \infty\}$.

For each $j = 1, 2, \cdots$ let A_j be a compact convex set such that $\mu(A_j') < j^{-1}$. Let $B_j = \cup_{i=1}^{j} A_i$, and let D_j be the convex hull of B_j. Then D_j is a compact convex set, $D_j \subset D_{j+1}$, and $\mu(D_j) \to 1$ as $j \to \infty$.

Choose a $\nu \in N_1$. Then $K(\nu) < \infty$; hence $\nu \ll \mu$, say, $(d\nu/d\mu)(v) = f(v)$, $0 \leqslant f < \infty$, and $K(\nu) = \int_V f(v) \log f(v) \mu(dv)$. It is plain that $a_j \equiv \nu(D_j) = \int_{D_j} f(v)\mu(dv) \to 1$ as $j \to \infty$. In the following assume j so large that $a_j > 0$, and let the probability measure ν_j be defined by $\nu_j(A) = \nu(A \cap D_j)/a_j$. It is easily checked that $K(\nu_j) \to K(\nu)$ as $j \to \infty$ and that $\int_V Y(v)\nu_j(dv) \to \int_V Y(v)\nu(dv)$; consequently $\nu_j \in N_1$ for all sufficiently large j. We observe next that $\nu_j(D_j) = 1$ and D_j is a compact convex set; hence $E(X_1|\nu_j)$ exists (Choquet (1969), II, page 115). Since $\nu \in N_1$ in this paragraph is arbitrary, it follows from (4.4) that

$$s(L_\theta) = \sup\{-K(\nu) : \nu \in N_1, E(X_1|\nu) \quad \text{exists}\}. \tag{4.5}$$

If $E(X_1|\nu)$ exists, say $E(X_1|\nu) = v$, then $\nu \in N_1$ if and only if $K(\nu) < \infty$ and $0 < \int_V Y(w)\nu(dw) = \int_V[\theta(w) - \theta(x) + \varepsilon]\nu(dw) = \theta(v) - \theta(x) + \varepsilon$, i.e., $v \in L_\theta$. It therefore follows from (4.5) that $s(L_\theta) \leqslant \sup\{-K(\nu) : E(X_1|\nu) \in L_\theta\}$; this last supremum equals $\sup\{\sigma(v) : v \in L_\theta\}$ by (3.7). The reverse inequality $s(L_\theta) \geqslant \sup\{\sigma(v) : v \in L_\theta\}$ is a consequence, e.g., of (4.1) and Theorem 2.3, and Lemma 4.2 is established. □

For each v in V let

$$g(v) = \limsup_{w \to v} \sigma(w), \qquad -\infty \leqslant g \leqslant 0. \tag{4.6}$$

As is well known, the concavity of σ (part (a)) and (4.6) imply that g is concave and upper semicontinuous. We note also that (4.1), (4.6), and the upper semicontinuity of s imply that

$$\sigma(v) \leqslant g(v) \leqslant s(v) \tag{4.7}$$

for all v.

Now choose $x \in V$, and positive constants m and δ, and let $t(x) = \max\{-m, g(x) + \delta\}$. It follows by replacing s with g and v with x in the second part of the proof of Theorem 3.2 that there exist θ in V^* and $\varepsilon > 0$ such that $g(v) < t(x)$ for $v \in L_\theta(x, \varepsilon)$. Hence $s(L_\theta(x, \varepsilon)) \leqslant t(x)$ by the first inequality in (4.7) and Lemma 4.2. Hence $s(x) \leqslant t(x)$ by the definition of $s(x)$; hence $s(x) \leqslant g(x)$; hence $s(x) = g(x)$ by (4.7). Since x is arbitrary, part (c) holds. Now let J be a nonempty open set. It is plain from (4.6) that, for each v in J, $g(v) \leqslant \sup\{\sigma(w) : w \in J\} = \sigma(J)$ say; hence, by part (c) and (4.6), $\operatorname{lan}(J) \leqslant \sigma(J)$. The reverse inequality $\operatorname{lan}(J) \geqslant \sigma(J)$ is evident from (4.1), so part (d) holds.

If $v \notin C$, $s(v) = -\infty$ by Theorem 2.4(a); hence $s(v) = \sigma(v)$ by (4.1). To establish part (e) it will therefore suffice to show that $s = \sigma$ on I. To this end, suppose I is nonempty. Then s is finite (and continuous) on I, by Theorem 2.4(c). Since I is open, it follows from part (c) that the set $G = \{w : \sigma(w) > -\infty\}$ is everywhere dense in I. Let z be a point in I. If V is k-dimensional there exist points $x_0, x_1, \cdots, x_k$ in I such that z is in the interior of the convex hull of $\{x_0, \cdots, x_k\}$. For each i let B_i be an open set containing x_i such that $B_i \subset I$ and such that if $y_i \in B_i$ for $i = 0, \cdots, k$ then z is in the convex hull of $\{y_0, \cdots, y_k\}$.

Now choose $y_i \in B_i \cap G$ for each i; then $\sigma(y_i) > -\infty$ for each i; hence $\sigma(z) > -\infty$ by part (a). Thus σ is concave and finite on the open convex set I; hence σ is continuous on I. It now follows from part (c) that $\sigma(v) = s(v)$ for $v \in I$; thus part (e) holds.

Suppose, finally, that V is one-dimensional. In view of part (e), it will suffice to show that if x is a boundary point of C then $s(x) = \sigma(x)$. The closed half spaces with x on their boundaries are $H_1 = \{v : v \leq x\}$ and $H_2 = \{v : v \geq x\}$. Suppose C is a subset of H_1. Then $s(H_1) = 0$ and $s(H_2) = \log p \leq 0$, where $p = \mu(\{x\})$. Hence $s(x) = \log p$ by Theorems 3.1, 3.2. If $p > 0$, the only measure in $M(x)$ which is absolutely continuous with respect to μ is the measure degenerate at x, and hence $\sigma(x) = \log p$; if $p = 0$, then $s(x) = -\infty$ and hence $s(x) = \sigma(x)$ by (4.1). This completes the proof of Theorem 3.3.

5. On the invariance of Lanford's theory under continuous linear transformations. We recall from Section 1 that V is a topological vector space equipped with a locally convex Hausdorff topology τ, that $\{X_n\}$ is a sequence of measurable transformations on $(\Omega, \mathcal{A}, P)$ into V such that $\{X_n(\omega)\}$ is an i.i.d. process with $PX_n^{-1} = \mu$, and that Assumptions 1, 2, and 3 hold. Now let V^0 be a real vector space and let V^0 be equipped with a topology τ^0 in which it is a locally convex Hausdorff topological vector space. Let ξ be a continuous linear function on V into V^0. For each $n = 1, 2, \cdots$ let $X_n^0(\omega) = \xi(X_n(\omega))$. In the following we refer to the entities V^0, τ^0, $\mu^0 = \mu\xi^{-1}$, and $\{X_n^0\}$ as the transformed framework.

THEOREM 5.1. *All conclusions of Sections* 1, 2, *and* 3 *hold in the transformed framework.*

PROOF. Let $\mathfrak{B}^0$ denote the Borel field in V^0. Since ξ is a measurable transformation of V into V^0, and since $\overline{X}_n^0 \equiv \xi(\overline{X}_n)$ and $Y_{m,n}^0 \equiv \xi(Y_{m,n})$, it is plain that the conditions stated in the paragraph containing (1.16), as well as Assumption 1, continue to hold with X_i, μ, and $\mathfrak{B}$ replaced by X_i^0, μ^0, and $\mathfrak{B}^0$. Let λ be a regular and convex-regular probability measure on V. It follows easily from the continuity of ξ that $\lambda^0 = \lambda\xi^{-1}$ is regular on V^0, and from the continuity and linearity of ξ that λ^0 is convex-regular. Since $\mu_n^0 \equiv \mu_n\xi^{-1}$ and each μ_n is regular and convex-regular, we conclude that Assumptions 2 and 3 also hold in the transformed framework. □

Let y denote the typical point in V^0, and let $s^0(y)$ be the point entropy at y in the transformed framework. According to Theorem 5.1, s^0 exists, and can be computed and used to estimate or bound large deviation probabilities concerning $\overline{X}_n^0$, by using the methods of Sections 1, 2, and 3 *in the transformed framework*. Of course, for any given $B \subset V^0$ we may, if we wish, express $\mu_n^0(B)$ as $\mu_n(A)$ with $A = \xi^{-1}(B)$, and estimate or bound $\mu_n(A)$ in terms of the original framework. The latter procedure amounts, roughly speaking, to replacing the actual entropy function s^0 on V^0 with another function, s_0 say, defined as follows: for each $y \in V^0$

$$s_0(y) = \sup\{s(v) : v \in \xi^{-1}\{y\}\} \tag{5.1}$$

if $\xi^{-1}\{y\}$ is nonempty and $s_0(y) = -\infty$ otherwise. We shall call s_0 the pullback entropy function induced by ξ. It is of some interest to enquire whether

$$s_0(y) \equiv s^0(y). \tag{5.2}$$

Example 5.1 at the end of this section shows that in general the pointwise invariance represented by (5.2) does not hold. A number of sufficient conditions for (5.2) are given in Theorems 5.2(f) and 5.3 below.

THEOREM 5.2.
(a) *The function* $s_0 : V^0 \to [-\infty, 0]$ *is concave.*
(b) *For all* $y \in V^0$, $s_0(y) \leqslant s^0(y)$.
(c) *For all* $y \in V^0$, $\limsup_{z \to y} s_0(z) = s^0(y)$.
(d) *For every nonempty open set* $B \subset V^0$, $\sup\{s_0(y) : y \in B\} = \sup\{s^0(y) : y \in B\}$.
(e) *If* V^0 *is finite-dimensional,* $\{y : s_0(y) \neq s^0(y)\}$ *is a subset of the boundary of the closed convex hull of the support of* μ^0.
(f) *If* $V^0 = R^1$ *then* $s_0(y) = s^0(y)$ *for all* $y \in R^1$.

PROOF. We shall establish only parts (b), (c), and (d). Let $B \subset V^0$ be a nonempty open convex set. The supremum of s^0 over B is the entropy of B, by Theorem 2.3^0. However, the entropy of B is also the entropy of $\xi^{-1}(B)$; the latter entropy equals the supremum of s_0 over B, by Theorem 2.3 and (5.1). Thus part (d) holds for open convex sets; that (d) holds for all open sets now follows from the local convexity of V^0. Part (c) is an immediate consequence of part (d) and Theorem 2.2^0. Part (b) is an immediate consequence of part (c). ☐

Theorems 3.3 and 5.2 suggest that the functions $s(v)$ and $\sigma(v)$ of the preceding sections are the actual and pullback entropies induced by some mapping of an i.i.d. process on some vector space into V. We now show that this suggestion is correct at least in the case when V is a Polish space. Let $\hat{V}$ be the space of finite signed measures on V, and let $\hat{V}$ be equipped with the weak topology. For each n let $\hat{X}_n(\omega)$ be the probability measure degenerate at $X_n(\omega)$. It follows from Section 7 (with the space Z of that section replaced by V) that $\{\hat{X}_n\}$ is a well-defined process on $\hat{V}$, that the conclusions of Sections 1, 2, and 3 hold for this process, and that the point entropy at a point ν in $\hat{V}$ is $-K(\nu)$ if ν is a probability measure and is $-\infty$ otherwise. Let $\hat{W}$ be the space of all ν such that $E(X_1|\nu)$ exists, and let $\eta(\nu) = E(X_1|\nu)$ on $\hat{W}$. Then η is a linear map of $\hat{W}$ onto V. Regard $\{\hat{X}_n\}$ as a process on $\hat{W}$. Since $\eta(\hat{X}_n(\omega)) \equiv X_n(\omega)$, it follows that $s(v)$ and $\sigma(v)$ are the entropy functions induced by η. Theorem 3.3 is, however, not a special case of Theorem 5.2, mainly because in general η is not continuous. For this same reason the present considerations applied to the framework of Example 3.2 do not yield a contradiction to (5.2).

As in preceding sections let V^* denote the set of real-valued τ-continuous linear functions θ on V. With V and V^* regarded as vector spaces in duality, let τ_w be the weak topology on V and τ_m^* the Mackey topology on V^*. Let ϕ denote the moment generating function of μ, i.e., $\phi(\theta) = \exp[c(\theta)] = \int_V \exp[\theta(v)]\,\mu(dv)$ for $\theta \in V^*$, $0 < \phi \leqslant \infty$. As in Sections 2 and 3, let C be the closed convex hull of the support of

μ in the τ-topology. For any real t let

$$A_t = \{v : s(v) \geqslant t\}, \qquad B_t = \xi(A_t). \tag{5.3}$$

We note that A_t is a τ-closed convex set; hence B_t is a convex set in V^0.

THEOREM 5.3. *Each of the following conditions is sufficient for* (5.2):

(i) *C is τ_w-compact.*

(ii) *V is a Banach space and* (1.15) *holds.*

(iii) *V is a reflexive Banach space* (e.g., $V = R^k = V^*$) *and $\phi(\theta) < \infty$ in a neighborhood of $\theta = 0$.*

(iv) *$\phi(\theta)$ is τ_m^*-continuous at $\theta = 0$.*

(v) *For each t, A_t is τ_w-compact.*

(vi) *For each t, B_t is τ^0-closed.*

PROOF. We observe first that both ϕ and c are proper convex functions on V^*, and that both are lower semicontinuous in the Mackey topology (and therefore in any topology compatible with the duality between V and V^*). That ϕ is convex is immediate from the convexity of the exponential function, and the convexity of c is a consequence of Hölder's inequality. To see that ϕ (and therefore c) is lower semicontinuous, choose and fix $\theta \in V^*$. Since μ is convex-regular, there exists a sequence $\{D_j : j = 1, 2, \cdots\}$ of compact convex sets of V such that $D_j \subset D_{j+1}$ for each j and $\mu(D_j) \to 1$ as $j \to \infty$ (cf. Section 4). Choose $\varepsilon > 0$ and for each j let $N_j(\varepsilon)$ be the set of all $\delta \in V^*$ such that $\sup\{|\delta(v) - \theta(v)| : v \in D_j\} < \varepsilon$. Since compact sets are weakly compact, D_j is a weakly compact convex set; hence $N_j(\varepsilon)$ is a Mackey-open neighborhood of θ. Hence

$$\begin{aligned} \liminf_{\delta\to\theta}\phi(\delta) &\geqslant \inf\{\phi(\delta) : \delta \in N_j(\varepsilon)\} \\ &\geqslant \inf\{\textstyle\int_{D_j} \exp[\delta(v)]\,\mu(dv) : \delta \in N_j(\varepsilon)\} \\ &\geqslant e^{-\varepsilon}\textstyle\int_{D_j} \exp[\theta(v)]\,\mu(dv). \end{aligned} \tag{5.4}$$

It follows from (5.4) by letting $\varepsilon \to 0$ and $j \to \infty$ that, whether $\phi(\theta) < \infty$ or $= +\infty$, $\liminf_{\delta\to\theta}\phi(\delta) \geqslant \phi(\theta)$; thus ϕ is lower semicontinuous on V^*.

We show next that each of the conditions (i) through (v) implies condition (vi). Since c is proper, convex, and lower semicontinuous, it follows from a theorem of Moreau (1966) that (iv) is equivalent to the condition that $\{v : c^*(v) \leqslant -t\}$ be weakly compact for each real t; Theorem 3.2 and (5.3) now imply that (iv) is equivalent to (v). To show that (v) implies (vi) let τ_w^0 be the weak topology on V^0 (determined by the given topology τ^0). Since ξ is a $\tau - \tau^0$-continuous linear function on V into V^0, it is easily seen that ξ is also $\tau_w - \tau_w^0$-continuous. Hence (v) and (5.3) imply that each B_t is τ_w^0-compact. Since τ_w^0 is a Hausdorff topology each B_t is weakly closed and therefore closed in V^0 and (vi) holds. We observe next from (5.3) and Theorem 2.4(a) that $A_t \subset C$. Since A_t is a closed convex set it is also weakly closed; it follows hence that (i) implies (v). It follows from Theorem 5.2 of Donsker and Varadhan (1976) (quoted at the end of Section 3) by Theorem 3.2 that (ii) implies that each A_t is compact and therefore weakly compact; thus (ii) implies (v). Suppose next that (iii) holds and let $\|\theta\|^* = \sup\{|\theta(v)| : v \in V, \|v\| \leqslant 1\}$ for $\theta \in V^*$. By the present hypothesis, there exists an $\varepsilon > 0$ such that $\|\theta\|^* < \varepsilon$ implies $\phi(\theta) < \infty$; since ϕ is convex, it follows (Rockafellar (1974), page 31) that ϕ is $\|\cdot\|^*$-continuous at $\theta = 0$. Since V is reflexive, the $\|\cdot\|^*$-topology on V^* is identical with the τ_m^*-topology, so that (iii) implies (iv). It is thus shown that (i) $\Rightarrow$ (v), (ii) $\Rightarrow$ (v) (and more), (iii) $\Rightarrow$ (iv), (iv) $\Leftrightarrow$ (v), and (v) $\Rightarrow$ (vi).

It will now suffice to show that (vi) implies (5.2). Choose a point $z \in V^0$, let m and ε be positive constants, and let $t = \max\{-m, s_0(z) + \varepsilon\}$. It follows from (5.1) and (5.3) then that the intersection of A_t and $\xi^{-1}\{z\}$ is empty; of course one or both of these sets might be empty. In any case $B = \xi(A_t)$ does not contain z. Since B is a closed convex set there exists an open convex set L^0 such that $z \in L^0$ and $L^0 \cap B = \varnothing$. Then $L = \xi^{-1}(L^0)$ is an open convex set such that the intersection of L and $\xi^{-1}(B)$ is empty. Since $A_t \subset \xi^{-1}(B)$ we have $s(v) < t$ for $v \in L$; hence $s(L) \leqslant t$ by Theorem 2.3. However, $s(L)$ is also the entropy of L^0; hence $s^0(z) \leqslant t$ by the definition of $s^0(z)$. Since m and ε in the definition of t are arbitrary it now follows from Theorem 5.2(b) that $s_0(z) = s^0(z)$. □

NOTE 1. It is no accident that some of the conditions of Theorem 5.3 resemble sufficient conditions (cf., e.g., Theorems 18, 18′ of Rockafellar (1974)), in order that certain convex optimization problems have solutions; indeed, the general theory of such problems can be used to obtain some necessary and sufficient conditions for (5.2), as follows. Let γ denote a real valued τ^0-continuous functional on V^0, and let Γ be the space of all such functionals γ. Choose and fix a point $z \in V^0$. Since $\gamma\xi \in V^*$ for each $\gamma \in \Gamma$,

$$F(\theta, \gamma | z) = c(\theta + \gamma\xi) - \gamma(z) \tag{5.5}$$

is a well-defined function on $V^* \times \Gamma$ into $(-\infty, +\infty]$. Let

$$f(\theta | z) = \inf\{F(\theta, \gamma | z) : \gamma \in \Gamma\}. \tag{5.6}$$

It follows easily from (5.5), (5.6) by Theorem 3.2^0 that

$$s^0(z) = f(0 | z). \tag{5.7}$$

It follows from (5.5) and (5.6) by application of Theorems 1 and 7 of Rockafellar (1974) that $f(\theta | z)$ is a convex function of θ, and that

$$s_0(z) = \liminf_{\theta \to 0} f(\theta | z) \tag{5.8}$$

in any compatible topology on V^*; we omit the verification. It follows from (5.7) and (5.8) that (5.2) holds if and only if $f(\theta | z)$ is lower semicontinuous at $\theta = 0$ for each $z \in V^0$. Another conclusion provided by the present methods is that (5.2) holds if and only if s_0 is an upper semicontinuous function on V^0; this conclusion is, however, available also from Theorem 5.2(c).

NOTE 2. We have seen in the proof of Theorem 5.3 that c is always a lower semicontinuous proper convex function. It follows (Rockafellar (1974), page 16) that $(-s)^* = c^{**} = c$. Thus there is always a one-to-one correspondence between cumulant generating functions c on V^* and point entropy functions s on V. Of course, different probability measures μ may have the same cumulant generating function; cf. Example 3.1(b).

We consider next an interesting special case. Let τ_1 be a topology on V weaker than τ but such that V remains a locally convex (Hausdorff) topological vector space under τ_1. The following corollary is then immediate from Theorems 5.1–5.3 with $V^0 = V$, $\tau^0 = \tau_1$, and ξ the identity:

COROLLARY 5.1. *All conclusions of Sections* 1, 2, *and* 3 *hold with* τ *replaced throughout by* τ_1. *With* s_1 *the point entropy function on* V *in the* τ_1 *topology,* $s_1(v) \geqslant s(v)$ *for all* v. *If any of the conditions* (i)–(vi) *of Theorem* 5.3 *is satisfied, then* $s_1(v) \equiv s(v)$.

We note that condition (vi) reduces in the present case to the condition that $s(v)$ be τ_1-upper semicontinuous on V. It follows hence from Theorem 5.2(c) that in the present case condition (vi) is necessary and sufficient for $s_1(v) \equiv s(v)$. Yet another sufficient condition for $s_1(v) \equiv s(v)$ is that the dual space of V under τ_1 be V^*; this is immediate from Theorem 3.2. This last condition is satisfied by an admissible τ_1 if and only if $\tau_w \subset \tau_1 \subset \tau$.

The following example shows that (5.2) can break down even in the finite-dimensional case (cf. Theorems 5.2 and 5.3).

EXAMPLE 5.1. Choose sequences $\{p_j : j = 0, 1, 2, \cdots\}$ and $\{a_j : j = 2, 3, \cdots\}$ of positive constants exactly as in Example 3.2. Let $V = R^3$, let $u_0 = (0, 0, 0)$, $u_1 = (1, 0, 0)$, and $u_j = (a_j, 1, j \cdot a_j)$ for all $j \geqslant 2$, and let μ be defined by $\mu(\{u_j\}) = p_j$ for $j = 0, 1, 2, \cdots$. Let $V^0 = R^2$, and let $\xi : V \to V^0$ be defined by $\xi(r_1, r_2, r_3) = (r_1, r_2)$. Let $z = (1, 0)$. We shall show that

$$s_0(z) = \log p_1, \qquad s^0(z) = \log(p_0 + p_1). \tag{5.9}$$

The moment generating function of μ evaluated at $\theta = (t_1, t_2, t_3) \in V^*$ is

$$\phi(\theta) = p_0 + p_1 \exp(t_1) + \Sigma_{j=2}^{\infty}\, p_j \exp[(t_1 + jt_3)a_j + t_2]. \tag{5.10}$$

It follows from (5.10) by (3.19) that here the natural parameter space is the union of Θ_1 and Θ_2, where $\Theta_1 = \{\theta : t_1 \leqslant 0, t_3 = 0\}$ and $\Theta_2 = \{\theta : t_3 < 0\}$. For each real r let $w_r = (1, 0, r)$, and let $h_i(r)$ denote the supremum of $\theta(w_r) - c(\theta)$ over Θ_i for $i = 1, 2$. Since $\theta(w_r) = t_1 + rt_3$, it follows from (5.10) by straightforward calculations that $h_1(r) = -\log(p_0 + p_1)$, and that $h_2(r) = -\log p_1$ if $r \geqslant 0$ and $h_2(r) = +\infty$ if $r < 0$. Hence $c^*(w_r) = \max\{h_1(r), h_2(r)\} = h_2(r)$; hence $s(w_r) = -h_2(r)$. Since $w_r = (z, r)$, it follows that $s_0(z) = \sup\{s(w_r) : -\infty < r < \infty\} = \log p_1$, and the first part of (5.9) is established. The second part of (5.9) is equivalent to (3.22) since the present μ^0 and z are the μ and v_1 of Example 3.2.

6. Some examples. In each example of this section V is a separable Banach space and it follows from Lemma 1.1 with $V_1 = V$ that Assumptions 1, 2, and 3 hold.

EXAMPLE 6.1 (the multinomial case). Suppose $V = R^k$, $k \geqslant 2$, and suppose μ assigns probabilities $p_1, \cdots, p_k$ respectively to the unit vectors $(1, 0, \cdots, 0), \cdots, (0, \cdots, 1)$, where $p_i > 0$, $\Sigma_1^k p_i = 1$. Here $V^* = R^k$, and for $\theta = (\theta_1, \cdots, \theta_k) \in V^*$, $c(\theta) = \log(\Sigma_1^k p_i \exp[\theta_i])$. Let V_1 be the set of all $v = (v_1, \cdots, v_k)$ in V with $v_i \geqslant 0$ for each i and $\Sigma_1^k v_i = 1$. It is known (Rockafellar (1974)) and easily verified that the Fenchel transform of c is

$$c^*(v) = \Sigma_{i=1}^k v_i \log(v_i/p_i) \tag{6.1}$$

for $v \in V_1$ and $c^*(v) = \infty$ on V_1'. We have

$$s(v) \equiv -c^*(v) \tag{6.2}$$

by Theorem 3.2.

In the present case $s(A)$ exists and equals $\operatorname{lan}(A)$ for every open $A \subset R^k$. To see this, note that $\mu_n(A) \leqslant \mu_n(\bar{A}) = \mu_n(\bar{A} \cap V_1)$ for every n, and that $\bar{A} \cap V_1$ is compact; hence

$$\limsup_{n\to\infty} a_n(A) \leqslant \operatorname{lan}(\bar{A} \cap V_1) \tag{6.3}$$

by Lemma 2.5. It is plain from (6.1), (6.2) and $0 \log 0 = 0$ that $s(v)$ is continuous on V_1; hence $\operatorname{lan}(\bar{A} \cap V_1) = \operatorname{lan}(A \cap V_1)$. Since $s = -\infty$ on V_1', it follows, as desired, that the upper bound in (6.3) equals $\operatorname{lan}(A)$.

The above conclusion is easily seen to be equivalent to the following: in the finite multinomial case, Sanov's theorem holds for every open set in the space of probability distributions of the single observation (cf. Section 7). An equivalent conclusion is obtained by ad hoc methods in Bahadur (1971). (There is a rectifiable error in Lemma 5.2 of Bahadur (1971).)

The special case $k = 2$, $p_1 = p_2 = \frac{1}{2}$, and $K = \{(\frac{1}{2}, \frac{1}{2})\}$ provides an example of a compact convex K for which $s(K)$ does not exist and (2.5) holds with equality.

EXAMPLE 6.2. Suppose that $V = R^k$, $k \geqslant 1$, and that μ is a given probability measure on the Borel sets of V. Of course we cannot compute the resulting point entropy function explicitly, but Theorems 2.1, 2.3, and 3.3(d) afford the following semi-explicit conclusion. For every J which is a finite union of open convex sets $s(J)$ exists and

$$s(J) = \operatorname{lan}(J) = \sup\{-K(\nu) : E(X_1|\nu) \in J\}. \tag{6.4}$$

Suppose henceforth that Θ includes a neighborhood of $\theta = 0$; in the present case this is equivalent to the existence of an $\varepsilon > 0$ such that $m = E_\mu(\exp[\varepsilon\|x\|]) < \infty$. It then follows from Theorem 3 of Bartfai (1978) that (6.4) holds for every open J. It also follows that the inequality (1.14) holds for every Borel set J. To see this, choose $r > 0$ and let $B = \{\|v\| : \|v\| \leqslant r\}$. Since $\|\bar{X}_n\| \leqslant n^{-1}\Sigma_1^n\|X_i\|$, it follows from Bernsteins's inequality that $\mu_n(B') \leqslant m^n \exp(-nr\varepsilon)$. Since $\mu_n(J) \leqslant \mu_n(\bar{J}) \leqslant 2\max\{\mu_n(\bar{J} \cap B), \mu_n(B')\}$, and since $\bar{J} \cap B$ is compact, it follows from Lemma 2.5 that the left-hand side of (1.14) is not greater than $\max\{\text{lan}(\bar{J}), \log m - r\varepsilon\}$; (1.14) now follows by letting $r \to \infty$. It would be interesting to know whether the present integrability assumption is essential to the conclusions of this paragraph. We note in this connection that in general $\text{lan}(J) \neq \text{lan}(\bar{J})$, even if (1.15) holds and J is open.

EXAMPLE 6.3. (Gaussian measure on a Hilbert space). Let V be a Hilbert space with inner product (v, w). Then $V^* = V$, and $\theta(v) = (\theta, v)$ for $\theta \in V^*$, $v \in V$. We assume that V is infinite dimensional but separable. Let $\{e_j : j = 1, 2, \cdots\}$ be a complete orthonormal system in V, let $\{\lambda_j : j = 1, 2, \cdots\}$ be positive constants such that $\Sigma_1^\infty \lambda_j < \infty$, and let T be the covariance operator defined by

$$Tv = \Sigma_1^\infty \lambda_j \cdot (v, e_j) \cdot e_j \tag{6.5}$$

for $v \in V$. Let μ be the probability measure on the Borel sets of V such that, for each $\theta \in V$, the distribution of (θ, v) under μ is normal with mean zero and variance $(T\theta, \theta)$. We have

$$(T\theta, \theta) = \Sigma_1^\infty \lambda_j \cdot (\theta, e_j)^2 = Q(\theta), \quad \text{say}. \tag{6.6}$$

It is known (cf., e.g., Kuo (1975)) that such a μ exists and is unique.

Choose v and θ in V, and let $H_\theta(v)$ be defined by (3.2). It follows from (3.3), or otherwise directly from the fact that $\theta(\bar{X}_n)$ is normal with mean zero and variance $Q(\theta)/n$, that $s(H_\theta(v))$ equals 0 if $\theta(v) \leqslant 0$ and equals $-[\theta(v)]^2/2Q(\theta)$ if $\theta(v) > 0$. Hence $s(v) = -\sup\{[\theta(v)]^2/2Q(\theta) : \theta \neq 0\}$ by (3.4) and Theorem 3.2. A straightforward calculation, which we omit, now shows that

$$s(v) = -\tfrac{1}{2}\Sigma_1^\infty \lambda_j^{-1} \cdot (v, e_j)^2 \qquad \text{for} \quad v \in V. \tag{6.7}$$

Let $V_0 = \{v : \Sigma_1^\infty \lambda_j^{-1} \cdot (v, e_j)^2 < \infty\}$; then V_0 is the range of the operator $T^{\frac{1}{2}}$. In view of (6.7), V_0 is precisely the set F of Section 2. The point entropy function is not continuous on V or even on V_0.

For each $\theta \in V^*$ let the probability measure μ_θ be defined as in Section 3 in the paragraph preceding Theorem 3.4. Since the covariance between the random variables $\theta(v)$ and $\delta(v)$ is $(T\theta, \delta)$ when μ obtains, an easy calculation shows that

$$E(X_1 | \mu_\theta) = T\theta. \tag{6.8}$$

It follows from (6.8) that W, the space of means, is the range of T, i.e., $W = \{v : \Sigma_1^\infty \lambda_j^{-2} \cdot (v, e_j)^2 < \infty\}$. It is interesting to note that W is a proper subset of V_0; consequently Corollary 3.3 does not apply to every v with $s(v) > -\infty$. We observe next that, under μ, $\{(v, e_j)/(\lambda_j)^{\frac{1}{2}} : j = 1, 2, \cdots\}$ is a sequence of independent $N(0, 1)$-variables. Hence $\{v : (v, e_j)/(\lambda_j)^{\frac{1}{2}} \geqslant 1$ for infinitely many $j\}$ is a set of

μ-measure 1. Consequently V_0 and W are both sets of μ-measure zero. It follows that if v is distributed in V according to some unknown one of the probability measures $\{\mu_\theta : \theta \in V\}$ then, for almost all v, the likelihood is a continuous but unbounded function of θ and the likelihood equation $E(X_1 | \mu_\theta) = v$ has no solution. It is thus seen that the inference methods considered in the final paragraph of Section 3 all fail in the present example.

According to a theorem of Itô (1970) there exists a point $x \in V$ and a closed subspace H such that S, the support of μ, is the set $x + H$. (Cf. Section 2.) Hence $C = S$. It follows from Lemma 2.3 by an application of the separating hyperplane theorem that, in the general case, $C \neq V$ if and only if there exists an open half-space of zero μ-measure. Since $Q(\theta) > 0$ for $\theta \neq 0$, no such half-space exists in the present case. We conclude that $S = C = V = I$. However, $F = V_0$ is an everywhere dense subset of V with empty interior; cf. Theorem 2.4.

It is known (cf., e.g., pages 159–165 of Kuo (1975) and Lemma 6.1 of Donsker and Varadhan (1976)) that (1.15) holds in the present example, and also in Examples 6.4 and 6.5 below. It follows (cf. Section 1) that in all three examples $s(J)$ exists and equals $\operatorname{lan}(J)$ for finite unions of open convex sets and also for all Borel sets J such that

$$\operatorname{lan}(J^\circ) = \operatorname{lan}(\bar{J}). \tag{6.9}$$

Let ε be given, $0 < \varepsilon < \infty$, and consider the particular set

$$J_\varepsilon = \{v : \|v\| \geqslant \varepsilon\}. \tag{6.10}$$

Then $J_\varepsilon^\circ = \{v : \|v\| > \varepsilon\}$. Choose $v \in J_\varepsilon \cap V_0$. Then, with δ a positive constant, $w = (1 + \delta)v$ is a point in J_ε°, and $s(w) = (1 + \delta)^2 s(v)$ by (6.7); since $s(v) > -\infty$ and δ is arbitrary, $\operatorname{lan}(J_\varepsilon^\circ) \geqslant s(v)$; since v is arbitrary, $\operatorname{lan}(J_\varepsilon^\circ) \geqslant \operatorname{lan}(J_\varepsilon \cap V_0) = \operatorname{lan}(J_\varepsilon)$; thus (6.9) holds for J_ε. We observe next that, since the point entropy is a concave function and since $s(0) = 0$, $\operatorname{lan}(J_\varepsilon) = \operatorname{lan}\{v : \|v\| = \varepsilon\} = \operatorname{lan}\{v : \|v\|^2 = \varepsilon^2, v \in V_0\}$. Since $\|v\|^2 = \Sigma_1^\infty (v, e_j)^2$, it follows from (6.7) by inspection that $\operatorname{lan}\{v : \|v\|^2 = \varepsilon^2, v \in V_0\}$ equals $-\varepsilon^2/(2\lambda_0)$ where $\lambda_0 = \max\{\lambda_j : j = 1, 2, \cdots\}$. We conclude that $s\{v : \|v\| \geqslant \varepsilon\}$ exists and equals $-\varepsilon^2/(2\lambda_0)$. This conclusion is available also from Theorem 7 of Sethuraman (1964).

EXAMPLE 6.4. (Wiener measure). Suppose V is the vector space of all continuous real valued functions v on the interval $[0, 1]$ of the real line with $v(0) = 0$. Let V be equipped with the topology of uniform convergence, τ say, and let $\mathfrak{B}$ be the resulting σ-algebra of Borel sets of V. Let μ be the standard Wiener measure on $\mathfrak{B}$, i.e., X_1 is the Gaussian process with mean zero and covariance function $\min\{t, u\}$ for $t, u \in [0, 1]$.

Let V_0 be the class of all functions v in V which are absolutely continuous and such that $\int_0^1 [v'(t)]^2 \, dt < \infty$. We shall show that

$$s(v) = -\tfrac{1}{2}\int_0^1 [v'(t)]^2 \, dt \tag{6.11}$$

for $v \in V_0$ and that $s(v) = -\infty$ for every $v \notin V_0$. It is known (cf., e.g., Kuo (1975), page 115) that every nonempty open set of V has positive μ-measure. Hence $S = V = C = I$, but $F = V_0$, F is of μ-measure zero, $\bar{F} = V$, and F° is empty.

For each $t \in [0, 1]$ let l_t be the point in V^* defined by $l_t(v) \equiv v(t)$, and let V_0^* be the span of the functionals $\{l_t : 0 \leq t \leq 1\}$. Let τ_0 denote the smallest topology in which each $\theta \in V_0^*$ is continuous. Since functionals in the vector space V_0^* separate the points of V, and since condition (ii) of Theorem 5.3 is satisfied, it follows from Corollary 5.1 that Theorem 3.2 with τ replaced by τ_0 will yield the common point entropy function in the τ and τ_0 topologies. Let θ be a point in V_0^*. Then there exist a positive integer k, constants $a_1, \cdots, a_k$ and points $t_1, \cdots, t_k$ in [0, 1] with $0 < t_1 < \cdots < t_k \leq 1$, such that $\theta(v) = \Sigma_1^k a_i v(t_i)$ for all v in V. With $b_i = \Sigma_{j=i}^k a_j$ and $t_0 = 0$ we have

$$\theta(v) = \Sigma_{i=1}^k b_i[v(t_i) - v(t_{i-1})], \tag{6.12}$$

so that θ is normally distributed with mean 0 and variance $d^2(\theta)$, where

$$d^2(\theta) = \Sigma_{i=1}^k b_i^2(t_i - t_{i-1}). \tag{6.13}$$

Hence $c(\theta) = \frac{1}{2}d^2(\theta)$. Choose a v in V. Holding k and $t_1, \cdots, t_k$ fixed, the maximum of $\theta(v) - \frac{1}{2}d^2(\theta)$ over all $a_1, \cdots, a_k$ is easily seen from (6.12) and (6.13) to equal

$$\tfrac{1}{2}\Sigma_{i=1}^k \frac{[v(t_i) - v(t_{i-1})]^2}{(t_i - t_{i-1})} = \tfrac{1}{2}f(k; t_1, \cdots, t_k), \quad \text{say.} \tag{6.14}$$

The supremum of f over all k and $t_1, \cdots, t_k$ equals $\int_0^1[v'(t)]^2\,dt$ if $v \in V_0$ and $+\infty$ if $v \in V_0'$ (Riesz and Nagy (1956), page 75). It now follows, as desired, that $s = -\infty$ on V_0' and that (6.11) holds on V_0.

Let J_ε be defined by (6.10). It then follows exactly as in Example 6.3 that $s(J_\varepsilon)$ exists and equals the supremum of $s(v)$ over all v in V_0 such that $\|v\| = \varepsilon$. It follows hence that $s(J_\varepsilon) = -\varepsilon^2/2$.

As noted previously by Donsker and Varadhan (1976) and others, if $\{X_n : n = 1, 2, \cdots\}$ is an i.i.d. process of centered Gaussian vectors then $\bar{X}_n/n^{1/2}$ has the same distribution as X_1; consequently, large deviation results concerning $\bar{X}_n$ are equivalent to certain results concerning X_1. It follows hence from the preceding paragraph that in the present example $n^{-1} \log P(\|X_1\| \geq \varepsilon n^{1/2}) \to -\varepsilon^2/2$ as $n \to \infty$. With t a parameter taking values in $[0, \infty)$ this last conclusion is equivalent to

$$\lim_{t\to\infty} t^{-1} \log P(\|X_1\| \geq \varepsilon t^{1/2}) = -\varepsilon^2/2. \tag{6.15}$$

Now let $\{Y(t) : 0 \leq t < \infty\}$ be a standard Wiener process, and for T a positive constant let $Z(T) = \sup\{|Y(t)| : 0 \leq t \leq T\}$. Since the processes $\{Y(tT)/T^{1/2} : 0 \leq t \leq 1\}$ and $\{X_1(t) : 0 \leq t \leq 1\}$ have the same distribution, the distribution of $Z(T)$ is the same as that of $T^{1/2}\|X_1\|$. It follows that (6.15) is equivalent to

$$\lim_{T\to\infty} T^{-1} \log P(Z(T) \geq \varepsilon T) = -\varepsilon^2/2 \tag{6.16}$$

and also to

(6.17) $$\lim_{T\to 0} T \log P(Z(T) \geqslant \varepsilon) = -\varepsilon^2/2.$$

The conclusions of this paragraph are special cases of results of Varadhan (1967), Marcus and Shepp (1972), Borell (1977), and others.

EXAMPLE 6.5. (Brownian bridge). Let V, τ, μ, and V_0 be the same as in Example 6.4 and let V_1 be the set of all $v \in V$ with $v(1) = 0$. Now let μ be replaced by the conditional μ-measure given that $v \in V_1$. It can be shown by the same method as is used in Example 6.4 that s is then given by (6.11) for $v \in V_1 \cap V_0$ and $s = -\infty$ elsewhere. It follows hence that with J_ε defined by (6.10) we have $s(J_\varepsilon) = -2\varepsilon^2$.

Let $\{Y(t) : 0 \leqslant t < \infty\}$ and $\{Z(T) : 0 < T < \infty\}$ be the processes defined in the last paragraph of Example 6.4. It then follows from $s(J_\varepsilon) = -2\varepsilon^2$ that $\lim_{T\to\infty} T^{-1} \log P(Z(T) \geqslant \varepsilon T | Y(T) = 0) = \lim_{T\to 0} T \log P(Z(T) \geqslant \varepsilon | Y(T) = 0) = -2\varepsilon^2$.

NOTE 1. The referee has kindly pointed out that in Examples 6.4 and 6.5 the point entropies in the τ and τ_0 topologies can be shown to be the same without using Theorem 5.3. Choose $\theta \in V^*$. There exists a function F of bounded variation on $[0, 1]$ such that $\theta(v) = \int_{[0,1]} v(t)F(dt)$ for all $v \in V$. There exists a sequence $\{F_j : j = 1, 2, \cdots\}$ such that each F_j is a step function on $[0, 1]$ with a finite number of steps, and such that $F_j \to F$ weakly. With $\theta_j(v) \equiv \int_{[0,1]} v(t)F_j(dt)$, $\{\theta_j\}$ is a sequence in V_0^* such that $\theta_j(v) \to \theta(v)$ for each v. If θ and θ_j are normally distributed with means 0 and variances $d^2(\theta)$ and $d^2(\theta_j)$ respectively, it follows from the continuity theorem for characteristic functions that $d^2(\theta_j) \to d^2(\theta)$. Hence $c_0^*(v) = \sup\{\delta(v) - \frac{1}{2}d^2(\delta) : \delta \in V_0^*\} \geqslant \theta(v) - \frac{1}{2}d^2(\theta)$ for each v. Since θ is arbitrary and $V_0^* \subset V^*$, $c_0^*(v) \equiv c^*(v)$.

NOTE 2. Examples 3.1(d), 6.3, 6.4, and 6.5 are special cases of a Gaussian measure on a Banach space. B. V. Rao and V. Mandrekar have kindly suggested to us that the general case can be treated as follows. Suppose that V is a separable Banach space with norm $\|\cdot\|$ and that μ is a centered Gaussian measure, i.e., for each $\theta \in V^*$ the random variable $\theta(v)$ is normally distributed with mean zero and variance $Q(\theta)$ say, where $0 \leqslant Q(\theta) < \infty$. Regard V^* as a subspace of $L_2 = L_2(V, \mathfrak{B}, \mu)$ and let M be the closure of V^* in L_2. For $f \in M$ let Rf denote the Bochner integral of $v \cdot f(v)$ with respect to μ; it is known that $\|v\|^2$ is μ-integrable; hence $\|v \cdot f(v)\| = \|v\| \cdot |f(v)|$ is μ-integrable and so Rf exists. With $(\cdot, \cdot)_2$ the L_2 inner product we have

(6.18) $$\begin{aligned}\theta(Rf) &= \theta\left(\int_V v \cdot f(v)\mu(dv)\right)\\ &= \int_V \theta(v \cdot f(v))\mu(dv)\\ &= \int_V \theta(v) \cdot f(v)\mu(dv)\\ &= (\theta, f)_2\end{aligned}$$

for all $\theta \in V^*, f \in M$. It follows from (6.18) that R is a linear one-to-one map of M onto a set $V_0 \subset V$. Define $(u, v)_0 = (R^{-1}u, R^{-1}v)_2$ and $\|u\|_0 = ((u, u)_0)^{\frac{1}{2}}$ for $u, v \in V_0$. Since M is a separable Hilbert space, so is the space V_0 with norm $\|\cdot\|_0$. (It is easily verified from (6.18) that, with elements of V_0 regarded as functionals on V^*, V_0 is the reproducing kernel Hilbert space of the covariance function $(\cdot, \cdot)_2$ on $V^* \times V^*$. Cf. Kuo (1975), Kuelbs (1976), and Mandrekar (1979) for various other properties of V_0. The present construction of V_0 is the one given in Mandrekar (1978).) It is shown in the following paragraph that

$$s(v) = -\tfrac{1}{2}\|v\|_0^2 \quad \text{for} \quad v \in V_0 \tag{6.19}$$

and

$$s(v) = -\infty \quad \text{for} \quad v \in V_0'. \tag{6.20}$$

It is plain from (6.19) and (6.20) that finding the point entropy function is always equivalent to finding the set V_0 and the norm $\|\cdot\|_0$ on V_0.

The cumulant generating function of μ is $c(\theta) = \frac{1}{2}Q(\theta) = \frac{1}{2}\|\theta\|_2^2$ for $\theta \in V^*$. To avoid trivialities we assume that $Q(\theta) > 0$ for at least one θ. Choose a point $x \in V_0$, say $x = Rf$. It then follows from (6.18) that $c^*(x)$ is the supremum of $(\theta, f)_2 - \frac{1}{2}\|\theta\|_2^2$ over V^*. Since $(g, f)_2 - \frac{1}{2}\|g\|_2^2 = Ug$ say is continuous on M and V^* is dense in M, $c^*(x)$ is the supremum of Ug for $g \in M$. For any r, $0 \leq r < \infty$, the maximum of Ug over the set $\{g : g \in M, \|g\|_2 = r\}$ is $r(f, f)_2 - \frac{1}{2}r^2$; hence $c^*(x) = \frac{1}{2}(f, f)_2 = \frac{1}{2}\|x\|_0^2$. Since x is arbitrary and $s = -c^*$, (6.19) is established. To establish (6.20) it will now suffice to show that if $x \in V$ and $c^*(x) = -s(x) < \infty$ then $x \in V_0$. By the definition of c^*, $\theta(x) - \frac{1}{2}\|\theta\|_2^2 \leq c^*(x)$ for all $\theta \in V^*$. It follows that $|\theta(x)| \leq a(x) < \infty$ for all $\theta \in V^*$ with $\|\theta\|_2 = 1$, where $a(x) = c^*(x) + \frac{1}{2}$, so that $\theta(x)$ is a bounded linear functional on V^*. There exists an extension of this functional to a bounded linear functional on M, say T. Since M is a Hilbert space there exists an $f \in M$ such that $Tg \equiv (g, f)_2$; in particular $T\theta = \theta(x) = (\theta, f)_2$ for $\theta \in V^*$. It now follows from (6.18) that $x = Rf$, so $x \in V_0$. The arguments just concluded show, incidentally, that $x \in V_0$ if and only if $\theta(x)$ is an L_2-bounded linear functional on V^*. It follows from (6.19), (6.20) by Theorems 2.1, 2.3, and Itô (1970) that V_0 is an everywhere dense subset of the support of μ.

For $\varepsilon > 0$ let J_ε be defined by (6.10). It follows from (6.19), (6.20) exactly as in Example 6.3 that $s(J_\varepsilon)$ exists in the present general case, and that

$$s(J_\varepsilon) = -\tfrac{1}{2}\inf\left\{\|v\|_0^2 : v \in V_0, \|v\| = \varepsilon\right\}. \tag{6.21}$$

7. The Sanov problem. Let $(Z, \mathcal{F}, p)$ be a probability space; here Z is a set of points z, $\mathcal{F}$ is a σ-algebra of subsets of Z, and p is a probability measure on $\mathcal{F}$. For any probability measure q on $\mathcal{F}$ let $\kappa(q) = \int_Z \log[(dq/dp)(z)]q(dz)$ if $q \ll p$ on $\mathcal{F}$, and let $\kappa(q) = +\infty$ otherwise. Let V_1 denote the set of all probability measures on $\mathcal{F}$. For any set $J \subset V_1$ define

$$\operatorname{san}(J) = \sup\{-\kappa(q) : q \in J\} \tag{7.1}$$

if J is nonempty and $\operatorname{san}(J) = -\infty$ otherwise.

Now let $Y_1, Y_2, \cdots$ be a sequence of independent observations, each Y_i taking values in Z according to the probability measure p. For each n and $Y_1, \cdots, Y_n$ let $\hat{p}_n$ denote the empirical measure based on the sample $Y_1, \cdots, Y_n$, i.e., $\hat{p}_n(A; Y_1, \cdots, Y_n) = (\#$ of indices i with $1 \leqslant i \leqslant n$ and $Y_1 \in A)/n$ for $A \in \mathcal{F}$. We shall say that Sanov's theorem holds for a set $J \subset V_1$ if $P(\hat{p}_n \in J)$ is well defined for each n and $\lim_{n\to\infty} n^{-1} \log P(\hat{p}_n \in J) = \operatorname{san}(J)$.

Suppose that V_1 is a topological space, and let $J \subset V_1$ be an open set. Assume that $\{\hat{p}_n \in J\}$ is a measurable event, and that the law of large numbers holds in the sense that, for each probability measure $q \in J$, $P(\hat{p}_n \in J | q) \to 1$ as $n \to \infty$; these assumptions are, of course, conditions on the topology of V_1. It then follows by an application of Lemma 6.1 in Bahadur (1971) that

$$\liminf_{n\to\infty} n^{-1} \log P(\hat{p}_n \in J | p) \geqslant \operatorname{san}(J). \tag{7.2}$$

Assume further that

$$\lim_{n\to\infty} n^{-1} \log P(\hat{p}_n \in J | p) \quad \text{exists} \tag{7.3}$$

for every J which is an open convex subset of V_1. For each $q \in V_1$, let $t(q)$ be the infimum of the limit in (7.3) over all such J which contain q. It then follows from (7.2) by the definition of san(J) that

$$-\kappa(q) \leqslant t(q) \tag{7.4}$$

for all $q \in V_1$. It is thus seen that Sanov's purely measure-theoretic point entropy is a lower bound for the topological point entropy generated by virtually any reasonable topology on V_1. We think that this is an important source of the difficulties and complications of the Sanov problem; as noted in Section 1, and as is evident from (7.2), the main problem is to find adequate asymptotic upper bounds for $a_n(J) = n^{-1} \log P(\hat{p}_n \in J | p)$.

Various sufficient conditions in order that Sanov's theorem hold for a given $J \subset V_1$ are given in Sanov (1957), Hoadley (1967), Sethuraman (1970), Borovkov (1967), Stone (1974), Sievers (1976), Donsker and Varadhan (1976), and Groeneboom et al. (1979). It is shown in the following paragraphs that if Z is a Polish space and J is a finite union of weakly open convex subsets of V_1 then Sanov's theorem holds for J. We believe that this sufficient condition supplements rather than supplants the other conditions cited. It is also shown, incidentally, that equality holds in (7.4) in the weak topology (and therefore in any larger topology which satisfies the conditions of the preceding paragraph).

It is assumed henceforth that Z is a Polish space and that $\mathcal{F}$ is the σ-algebra of Borel sets of Z. Let V be the vector space of finite signed measures on $\mathcal{F}$. Let $\mathcal{C}$ be the class of all functions $f : Z \to R^1$ such that f is continuous and bounded. For each $f \in \mathcal{C}$ let l_f be the linear functional on V defined by

$$l_f(v) = \int_Z f(z) v(dz). \tag{7.5}$$

Let τ be the smallest topology in which l_f is continuous for each $f \in \mathcal{C}$. Since each point in V is the difference of two finite nonnegative measures it follows easily from Billingsley (1968), page 9, that the points of V are separated by functionals in

$\{l_f : f \in \mathcal{C}\}$. Since $\{l_f : f \in \mathcal{C}\}$ is a vector space, it follows that τ is an admissible topology on V, and that $V^* = \{l_f : f \in \mathcal{C}\}$. It should be noted that now the relative topology on V_1 is that of weak convergence of probability measures, and (Parthasarathy (1967), pages 45–46) that V_1 is a Polish space.

We observe next that V_1 is a convex and closed subset of V. The convexity of V_1 is obvious; the fact that V_1 is closed in V is well known and easily established by standard arguments (cf., e.g., Billingsley (1968)).

For each z in Z let δ_z denote the probability measure degenerate at z, and let $T : Z \to V$ be the map which takes z into δ_z. It is readily seen that T is a continuous map; consequently T is an $\mathcal{F}$-$\mathcal{B}$-measurable transformation of Z into V, where $\mathcal{B}$ is, as usual, the Borel field in V. Hence

$$\mu = pT^{-1} \tag{7.6}$$

is well defined on $\mathcal{B}$. Clearly, $\mu(V_1) = 1$.

Let $(\Omega, \mathcal{Q}, P)$ be a probability space and for each n let $Y_n : \Omega \to Z$ be an $\mathcal{Q}$-$\mathcal{F}$-measurable transformation such that, under P, $\{Y_n(\omega) : n = 1, 2, \cdots\}$ is an i.i.d. process with $PY_n^{-1} = p$. For each n let $X_n(\omega) = TY_n(\omega)$. It then follows from the preceding discussion first that $(\Omega, \mathcal{Q}, P)$ and $\{X_n(\omega) : n = 1, 2, \cdots\}$ satisfy all the conditions listed in the paragraph containing (1.16). Since $\overline{X}_n \equiv \hat{p}_n$, it then follows from the preceding discussion by Lemma 1.1 that $\{\omega : \hat{p}_n \in J\}$ is $\mathcal{Q}$-measurable for every $J \in \mathcal{B}$ and every n, and that if J is a relatively open convex subset of V_1 then (7.3) holds and the limit equals lan(J). Since V_1 is a closed convex set with $\mu(V_1) = 1$, it also follows (cf. Lemma 2.3, Theorem 2.4(a)) that for Lanford's point entropy s we have

$$s(v) = -\infty \quad \text{on} \quad V_1'. \tag{7.7}$$

To establish Sanov's theorem for finite unions of relatively open convex sets of V_1 it remains to show that lan(A) = san(A) for such sets A. We shall show that in fact

$$s(q) = -\kappa(q) \qquad \text{for all} \qquad q \in V_1. \tag{7.8}$$

Let $D = \{\delta_z : z \in Z\}$. It is known (Parthasarathy (1967), page 42) that D is a sequentially closed subset of V_1. Since V_1 is metrizable and closed, D is a closed and therefore Borel measurable subset of V. Let $\mathcal{D}$ be the σ-algebra of Borel sets of D, i.e., $\mathcal{D} = \{D \cap B : B \in \mathcal{B}\}$. It is known that T is a homeomorphism between the topological spaces Z and D (Parthasarathy (1967), page 42). It follows that the measure spaces $(Z, \mathcal{F}, p)$ and $(D, \mathcal{D}, \mu)$ are isomorphic whenever p and μ are related by (7.6).

Choose a $\theta \in V^*$, say $\theta = l_f$. With ϕ the moment generating function of μ,

$$\begin{aligned} \phi(\theta) &= \textstyle\int_V \exp[\theta(v)]\,\mu(dv) \\ &= \textstyle\int_D \exp[\theta(v)]\,\mu(dv) \\ &= \textstyle\int_D \exp[l_f(v)]\,\mu(dv) \\ &= \textstyle\int_Z \exp[f(z)]\,p(dz), \end{aligned} \tag{7.9}$$

since $\mu(D) = 1$ and since $l_f(\delta_z) = f(z)$ by (7.5). It follows from (7.5) and (7.9) that

$$(7.10) \qquad c^*(v) = \sup\{\textstyle\int_Z f(z)v(dz) - \log \int_Z \exp[f(z)]p(dz) : f \in \mathcal{C}\}$$

for all v in V. It follows from (7.10) by Lemma 2.1 of Donsker and Varadhan (1975) that

$$(7.11) \qquad c^*(q) = \kappa(q) \quad \text{for all} \quad q \in V_1.$$

It follows from (7.11) by Theorem 3.2, as desired, that (7.8) holds.

Let q be a measure in V_1 and ν a probability measure on $\mathfrak{B}$. It is readily seen that $E(X_1|\nu) = q$ and $\nu \ll \mu$ if and only if $q \ll p$ and $\nu = qT^{-1}$; and that, in this case, $K(\nu) = \kappa(q)$. It follows hence that

$$(7.12) \qquad \sigma(q) = -\kappa(q) \qquad \text{for all} \quad q \in V_1.$$

In view of (7.7), (7.12), and Theorem 3.3, (7.8) is equivalent to the statement that κ is a lower semicontinuous function on V_1. We note also that the easily established (7.12) and Corollary 3.3 imply that $s = -\kappa$ on W; W is, however, a proper subset of V_1 in the general case.

EXAMPLE 7.1. (Sanov (1957)). This example shows that in general Sanov's theorem does not hold for all relatively open sets of V_1; the example therefore also shows that in general (1.8) and (1.13) do not hold for all open sets of V. Assume that p is a nonatomic probability measure on the Borel field $\mathfrak{F}$ of the Polish space Z. Let $A = \{v : v \in V, s(v) < -1\}$, and $J = A \cap V_1$. Then, by the upper semicontinuity of s, A is an open subset of V. If q is a purely atomic probability measure on $\mathfrak{F}$ then q is not dominated by p; hence $\kappa(q) = +\infty$, and hence $q \in J$ by (7.8). It follows that $\hat{p}_n \in J$ for all ω and n; hence $s(J) = 0$. However, $\text{lan}(J) = \text{san}(J) \leqslant -1$ by (7.8) and the definition of J. The present assumption that p is nonatomic implies, of course, that Z is an uncountably infinite set. It would be interesting to know whether Sanov's theorem can fail for open sets even when Z is countably infinite (cf. Example 6.1).

In view of (7.10) and (7.11), an important theorem of Donsker and Varadhan ((1976), Theorem 4.5) may be stated as follows: under the present assumptions that Z is a Polish space and that the topology on V_1 is that of weak convergence, (7.2) holds for every open $J \subset V_1$, and

$$(7.13) \qquad \limsup_{n\to\infty} n^{-1} \log P(\hat{p}_n \in K|p) \leqslant \text{san}(K)$$

holds for every closed $K \subset V_1$. It follows from this theorem that if $A \subset V_1$ is Borel measurable, and if $\text{san}(A^\circ) = \text{san}(\bar{A})$, then Sanov's theorem holds for A. An even weaker sufficient condition for Sanov's theorem is given in Theorem 3.1 of Groeneboom et al. (1979).

The bound (7.13) is, of course, a special case of (1.14). As noted in Section 1, it seems difficult to establish (1.14) in the general case. It might be added (cf. Example 6.2) that at present we know of no counterexamples to (1.14). In particular, (1.14) holds with equality for the sets A and J of Example 7.1.

The following is a proof of the bound (7.13) of Donsker and Varadhan (1976) based on the present methods. We have seen that with τ the weak topology on the present V, the present framework V, τ, μ, and $\{X_n\}$ satisfies Assumptions 1, 2, and 3, that $\bar{X}_n \equiv \hat{p}_n$, and that (7.7) holds. Let $\|f\| = \sup\{|f(z)| : z \in Z\}$ for functions in $\mathcal{C}$, let V^0 be the dual space of the resulting Banach space $\mathcal{C}$, and let τ^0 be the weak*-topology on V^0. For each v in V let $\xi(v)$ be the functional on $\mathcal{C}$ defined by $[\xi(v)](f) \equiv l_f(v)$ where $l_f(v)$ is given by (7.5). Then ξ is a continuous linear function on V into V^0. It follows from (7.7), (7.8) by Lemma 2.3 of Groeneboom et al. (1979) that condition (v) of Theorem 5.3 is satisfied; hence (5.2) holds for the actual and pullback entropies induced by ξ. With $R = \xi(V)$, ξ is a $1 - 1$ map of V onto R. It follows hence from (5.2) and (7.7) that, with $R_1 = \xi(V_1)$, the actual entropy on V^0 is given by

$$s^0(\xi(q)) = s(q) \quad \text{for} \quad q \in V_1 \tag{7.14}$$

and

$$s^0(y) = -\infty \quad \text{for} \quad y \in R_1'. \tag{7.15}$$

We observe next that ξ is a homeomorphism between the spaces V_1 and R_1 in their relative topologies. Consequently, given a closed set K in V_1 there exists a closed $L \subset V^0$ such that

$$\xi(K) = L \cap R_1. \tag{7.16}$$

Since R_1 is a subset of $B_1 = \{y : |y(f)| \leqslant 1 \text{ for all } f \text{ with } \|f\| \leqslant 1\}$, and since B_1 is compact, we may suppose that the set L in (7.16) is compact. Although $\xi(K)$ and R_1 might not be Borel sets, it follows from (7.16) that $\{\bar{X}_n \in K\}$ and $\{\xi(\bar{X}_n) \in L\}$ are the same $\mathcal{Q}$-measurable sets in Ω, so $\mu_n(K) \equiv \mu_n^0(L)$. Hence

$$\begin{aligned}
\limsup_{n\to\infty} a_n(K) &= \limsup_{n\to\infty} a_n^0(L) \\
&\leqslant \operatorname{lan}^0(L) && \text{by Lemma } 2.5^0 \\
&= \operatorname{lan}^0(L \cap R_1) && \text{by (7.15)} \\
&= \operatorname{lan}^0(\xi(K)) && \text{by (7.16)} \\
&= \operatorname{lan}(K) && \text{by (7.14)} \\
&= \operatorname{san}(K) && \text{by (7.8).}
\end{aligned}$$

NOTE. In this section we have regarded the Sanov problem as a special case of the large deviations problem for the sample mean. In Groeneboom et al. (1979) the authors travel in the opposite direction; they first study the Sanov problem, and then obtain conclusions concerning the sample mean and many other statistics by regarding these statistics as functions of the empirical probability measure. Adequate discussion of this important work cannot be undertaken here and the following remarks must suffice: the main conclusions of Groeneboom et al. concerning the Sanov problem are stronger or more useful than ours (partly because their topology on V_1 is much stronger than the weak topology), and their applications include certain results which are very close to our Theorems 2.1 and 2.3.

APPENDIX

PROOF OF LEMMA 1.1. Let V be a real vector space, τ an admissible topology on V (cf. Section 1), and let $\mathfrak{B}$ be the σ-algebra generated by the τ-open sets of V. Let λ be a probability measure on $\mathfrak{B}$. λ is said to be *inner regular* if $\lambda(B) = \sup\{\lambda(K) : K \text{ compact}, K \subset B\}$ for all $B \in \mathfrak{B}$. An inner regular measure is evidently regular in the sense of Section 1. It is well known that every probability measure on the Borel sets of a Polish space is inner regular. λ is said to be *convex-tight* if for every $\varepsilon > 0$ there exists a convex compact K_ε such that $\lambda(K_\varepsilon') < \varepsilon$. Since V is an open convex set it is clear that convex-regularity in the sense of Section 1 implies convex-tightness. The following is a useful partial converse.

PROPOSITION 1. *If λ is regular and convex-tight then λ is convex-regular.*

PROOF. Let J be an open convex set. Choose $\varepsilon > 0$. There exists a compact $K_1 \subset J$ with $\lambda(K_1) \geqslant \lambda(J) - \varepsilon$, and a compact convex K_2 with $\lambda(K_2') < \varepsilon$. Let $K_3 = K_1 \cap K_2$. Then $\lambda(K_3) \geqslant \lambda(K_1) - \lambda(K_2') \geqslant \lambda(J) - 2\varepsilon$. Since ε is arbitrary, it will suffice to show that there exists a compact convex set $\tilde{G}$ such that $K_3 \subset \tilde{G} \subset J$.

Since J is open and V is a regular space, it follows from local convexity that for each $x \in J$ there exists an open convex neighborhood of x, say $B(x)$, such that $\overline{B(x)} \subset J$. Since $\{B(x) : x \in K_1\}$ is an open covering of K_1, there exist open convex sets $B_1, \cdots, B_k$ such that $\overline{B}_i \subset J$ for each i and $K_1 \subset \cup_1^k B_i$. For each i, let $C_i = K_1 \cap B_i$, let D_i be the closed convex hull of C_i, and let $F_i = K_2 \cap D_i$. Then each F_i is a compact convex set. It follows that with $G = \cup_1^k F_i$, and $\tilde{G}$ the convex hull of G, $\tilde{G}$ is a compact convex set. Now, $K_1 = \cup_1^k (K_1 \cap B_i) = \cup_1^k C_i \subset \cup_1^k D_i$; hence $K_3 = K_1 \cap K_2 \subset \cup_1^k F_i = G$; hence $K_3 \subset \tilde{G}$. We observe next that $C_i \subset \overline{B}_i$, and $\overline{B}_i$ is closed and convex; hence $D_i \subset \overline{B}_i \subset J$; hence $F_i \subset J$; hence $G \subset J$. Since J is convex, $\tilde{G} \subset J$. □

Now let V_1 be a closed convex subset of V such that V_1 is a Polish space in its relative topology.

PROPOSITION 2. *If $\lambda(V_1) = 1$ then λ is regular and convex-regular.*

PROOF. Let $\mathfrak{B}_1$ be the σ-algebra of Borel sets of V_1, i.e., $\mathfrak{B}_1 = \{V_1 \cap B : B \in \mathfrak{B}\}$, and let λ_1 denote the restriction of λ to $\mathfrak{B}_1$. Since V_1 is Polish it follows that λ_1 is inner regular on V_1. It follows hence that λ is inner regular on V; in particular, λ is regular. In view of Proposition 1 it will now suffice to show that λ is convex-tight. Choose $\varepsilon > 0$. Since λ is regular there exists a compact K_1 such that $\lambda(K_1) > \lambda(V) - \varepsilon = 1 - \varepsilon$. We may assume that $K_1 \subset V_1$. Let K_2 be the closed convex hull of K_1. Since $\lambda(K_2) > 1 - \varepsilon$, it will suffice to show that K_2 is compact. Since K_2 is closed, and $K_2 \subset V_1$, K_2 is complete. We observe next that since K_1 is compact, $\tilde{K}_1$ is totally bounded (Choquet (1969), I, page 358). Since V is a regular space, K_2 also is totally bounded. □

To establish Lemma 1.1, let $(\Omega, \mathcal{Q}, P)$ be a probability space and $\{X_n(\omega) : n = 1, 2, \cdots\}$ a sequence of functions on Ω into V such that all the conditions listed in the paragraph containing (1.16) are satisfied.

It is known and easily verified that if C and D are second-countable topological spaces, and $\mathcal{C}$ and $\mathfrak{D}$ are the Borel fields of sets of C and D respectively, then the Borel field of the space $C \times D$ equipped with the product topology is exactly the product σ-algebra $\mathcal{C} \times \mathfrak{D}$. It follows hence by standard arguments that if V is second-countable (e.g., if V is Polish) then Assumption 1 is automatically fulfilled. Verification that Assumption 1 holds under the hypotheses of Lemma 1.1 is a trivial extension of these standard arguments and so is omitted. As noted in Section 1, Assumption 1 implies that each μ_n is a well-defined probability measure on $\mathfrak{B}$. The hypothesis $X_n \in V_1$ for all n implies $\bar{X}_n \in V_1$ and hence $\mu_n(V_1) = 1$ for all n. It now follows from Proposition 2 with $\lambda = \mu_n$ that each μ_n is regular and convex-regular; thus Assumptions 2 and 3 also hold.

Since V_1 is a closed convex set with $\mu(V_1) = 1$, (7.7) holds and the last part of Lemma 1.1 follows by straightforward modifications of some of the arguments of Sections 1 and 2. This completes the proof of Lemma 1.1.

Acknowledgment. The authors are very grateful to the referee for his careful reading of the paper and for many helpful comments and suggestions.

REFERENCES

[1] Abrahamson, I. G. (1965). On the stochastic comparison of tests of hypotheses. Doctoral dissertation, Univ. of Chicago.

[2] Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324.

[3] Bahadur, R. R. (1971). *Some Limit Theorems in Statistics*. SIAM, Philadelphia.

[4] Bahadur, R. R. and Ranga Rao, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 43–54.

[5] Bahadur, R. R. and Zabell, S. L. (1979). A lower bound for large deviation probabilities. (Manuscript in preparation.)

[6] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, New York.

[7] Bartfai, P. (1978). Large deviations of the sample mean in Euclidean spaces. Mimeograph Series No. 78–13, Statist. Depart., Purdue Univ.

[8] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[9] Borell, C. (1977). Tail probabilities in Gauss space. Unpublished manuscript.

[10] Borovkov, A. A. (1967). Boundary-value problems for random walks and large deviations in function spaces. *Theor. Probability Appl.* **12** 575–595.

[11] Borovkov, A. A. and Rogozin, B. A. (1965). On the multidimensional central limit theorem. *Theor. Probability Appl.* **10** 55–62.

[12] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[13] Choquet, G. (1969). *Lectures on Analysis*, **1** and **2**. Benjamin, New York.

[14] Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Sci. Indust.* **736** 5–23.

[15] Donsker, M. D. and Varadhan, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time-I. *Comm. Pure Appl. Math.* **27** 1–47.

[16] 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.

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

[18] GUPTA, J. C. (1972). Probabilities of medium and large deviations with statistical applications. Doctoral dissertation, Univ. of Chicago.

[19] HO, NGUYEN VAN (1974). Asymptotic efficiency in the Bahadur sense for the signed rank tests. *Proc. Prague Symp. on Asymptotic Statistics* (September 1973) **2** 127–156.

[20] HOADLEY, A. B. (1967). On the probability of large deviations of functions of several empirical cdf's. *Ann. Math. Statist.* **38** 360–382.

[21] HOEFFDING, W. (1965a). On probabilities of large deviations. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 203–219. Univ. of California Press.

[22] HOEFFDING, W. (1965b). Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.* **36** 369–408.

[23] IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

[24] ITÔ, K. (1970). The topological support of Gauss measure on Hilbert space. *Nagoya Math. J.* **38** 181–183.

[25] KUELBS, J. (1976). The law of the iterated logarithm and related strong convergence theorems for Banach space valued random variables. *Lecture Notes in Mathematics* **539** 224–314. Springer-Verlag, New York.

[26] KUO, H.-H. (1975). Gaussian measures in Banach spaces. *Lecture Notes in Mathematics* **463**. Springer-Verlag, New York.

[27] LANFORD, O. E. (1971). Entropy and equilibrium states in classical statistical mechanics. In *Statistical Mechanics and Mathematical Problems*. *Lecture Notes in Physics* **20** 1–113. Springer-Verlag, Berlin.

[28] MANDREKAR, V. (1979). *Gaussian Processes and Their Markov Property*. Almquist and Wicksell, Uppsala. (To appear.)

[29] MARCUS, M. B. and SHEPP, L. A. (1972). Sample behavior of Gaussian processes. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 423–439. Univ. of California Press.

[30] MOREAU, J. J. (1966). *Fonctionelles convexes*. Séminaire sur les équations aux dérivées partielles. Collège de France, Paris.

[31] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

[32] PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer-Verlag, New York.

[33] RIESZ, F. and NAGY, B. SZ.- (1955). *Functional Analysis*. Ungar, New York.

[34] ROCKAFELLAR, R. T. (1974). *Conjugate Duality and Optimization*. SIAM, Philadelphia.

[35] RUELLE, D. (1965). Correlation functionals. *J. Math. Physics* **6** 201–220.

[36] SANOV, I. N. (1957). On the probability of large deviations of random variables (in Russian). *Mat. Sb.* **42** 11–44. (English translation in *Selected Transl. Math. Statist. Prob.* (1961) **1** 213–244.)

[37] SETHURAMAN, J. (1964). On the probability of large deviations of families of sample means. *Ann. Math. Statist.* **35** 1304–1316. (Corrections (1970), *Ann. Math. Statist.* **41** 1376–1380.)

[38] SETHURAMAN, J. (1965). On the probability of large deviations of the mean for random variables in $D[0, 1]$. *Ann. Math. Statist.* **36** 280–285.

[39] SETHURAMAN, J. (1970). Probabilities of deviations. In *Essays in Probability and Statistics* 655–672. Univ. North Carolina Press, Chapel Hill.

[40] SIEVERS, G. L. (1975). Multivariate probabilities of large deviations. *Ann. Statist.* **3** 897–905.

[41] SIEVERS, G. L. (1976). Probabilities of large deviations for empirical measures. *Ann. Statist.* **4** 766–770.

[42] STONE, M. (1974). Large deviations of empirical probability measures. *Ann. Statist.* **2** 362–366.

[43] VARADHAN, S. R. S. (1967). Diffusion processes in a small time interval. *Comm. Pure Appl. Math.*, **20** 659–685.

[44] WOODWORTH, G. G. (1970). Large deviations and Bahadur efficiency of linear rank statistics. *Ann. Math. Statist.* **41** 251–284.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637