*The Annals of Mathematical Statistics*
1971, Vol. 42, No. 5, 1706–1716

# HIDE AND SEEK, DATA STORAGE, AND ENTROPY[1]

By Robert J. McEliece and Edward C. Posner
*California Institute of Technology*

**1. Introduction.** In this paper we will study the relationship between games of search and the optimum storage of information. In Section 2 we shall treat the case of finite sets, and in Section 3 a generalization to compact metric spaces. The result is a synthesis of the epsilon entropy theory of approximation (Lorentz (1966)), with the theory of data transmission and compression (Posner and Rodemich (1971)).

Let $X$ be a set with $n = |X|$ elements, and let $\mathscr{S} = \{S_1, S, \cdots, S_m\}$ be a finite collection of subsets of $X$, with $\cup S_j = X$. Regard the $x \in X$ as "data points," and the $S_j$ as "subsets of allowed uncertainty," such that when a data point is selected, one is not interested in exactly which $x$ it is, but rather in knowing an $S_j$ (there may be more than one) in which it lies. The class $\mathscr{S}$ is regarded as chosen by the experimenter.

Under these assumptions, if $m'$ is the fewest number of the $S_j$'s which are needed to cover $X$, then at least $\lceil \log_2 m' \rceil$ bits are needed in order to identify an $S_j$ in which an arbitrary $x$ lies. The method of specification to achieve $\lceil \log_2 m' \rceil$ is, of course, to merely specify the index $j$ of the set $S_j$ which contains the given data point, and one specifies $j$ by using a $\lceil \log_2 m' \rceil$-tuple of zeros and ones. This problem is considered in Balinsky (1968) pages 214–221.

However, if $N$ data points can be stored before it is attempted to specify a sequence of $N$ $S_j$'s, a saving may be possible. Let $X^N$ be the cartesian $N$th power of $X$, and let $\mathscr{S}^N$ be the class of subsets of $X^N$ of the form $S_{i_1} \times \cdots \times S_{i_N}$. Here $\lceil \log_2 M \rceil$ bits are needed to specify the sequence of sets corresponding to an unknown sequence of $N$ data points, where $M$ is the fewest number of sets from $\mathscr{S}^N$ needed to cover $X^N$. Thus $1/N\lceil \log_2 M \rceil$ can be interpreted as the number of bits per sample necessary to specify an $S_j$ when a "block code" of (constant) length $N$ is used.

Thus we are led to several definitions, which generalize those in Posner and Rodemich (1971); here is also found more information-theory, as well as a list of prior references for some of these concepts. The (one shot) *$\mathscr{S}$-entropy* $H_{\mathscr{S}}(X)$ is defined as

$$H_{\mathscr{S}}(X) = \min_{\mathscr{T} \subset \mathscr{S}} \log |\mathscr{T}|$$

where the minimization is taken over those subsets of $\mathscr{S}$ which cover $X$. (We

Received May 19, 1970.

[1] This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under Contract No. NAS 7–100, sponsored by the National Aeronautics and Space Administration. We are indebted to the referee for pointing out the reference to Integer Linear Programming.

disregard the rounding-off of the logarithm, and the base of the logarithm.) The limit, which can be shown to exist using an obvious subadditivity,

$$I_{\mathscr{S}}(X) = \lim_{N\to\infty} \frac{1}{N} H_{\mathscr{S}^N}(X^N),$$

the *absolute $\mathscr{S}$-entropy* of $X$, can be interpreted as the minimum number of bits per sample needed when arbitrarily long block codes are used. Note that

$$I_{\mathscr{S}}(X) \leqq H_{\mathscr{S}}(X).$$

(This concept implicitly occurs in graph theory (Berge (1962) Chapter 4), where $X$ is the set of vertices of a finite graph, and $\mathscr{S}$ is the set of *talons*; a talon is a vertex together with all the vertices to which it is connected. We shall have more to say about this later.)

Now let us provide $X$ with a probability measure $P$, and suppose the $S_j$ are $P$-measurable. Define the (one-shot) $\mathscr{S}$; $P$ *entropy* of $X$, $H_{\mathscr{S};P}(X)$, as the minimum Shannon entropy $H(\mathscr{V})$ of any partition of $X$ by subsets of the $S_j$; i.e., of any *$\mathscr{S}$-partition*: $H(\mathscr{V}) = \Sigma P(T_k) \log (1/P(T_k))$, where $X = \bigcup_k T_k$ and $\mathscr{V} = \{T_k\}$ is the partition. $H_{\mathscr{S};P}(X)$ is the minimum expected number of bits per sample necessary to specify the source $X$ using variable-length coding, when $X$ is sampled according to the probability distribution $P$. Again, the sets in $\mathscr{S}$ are the sets of allowed uncertainty, and round-off is ignored. See Posner and Rodemich (1971), for a full information-theoretic justification.

The *absolute $\mathscr{S}$;$P$ entropy* $I_{\mathscr{S};P}(X)$ is likewise defined as

$$I_{\mathscr{S};P}(X) = \lim_{N\to\infty} \frac{1}{N} H_{\mathscr{S}^N;P^N}(X^N),$$

where $P^N$ is the produce measure on $X^N$. Then $I_{\mathscr{S};P}(X)$ is the minimum expected number of bits per sample necessary to describe $X$ to within the uncertainty $\mathscr{S}$ when arbitrarily many samples can be stored and optimum variable-length coding is used.

We now change the subject and define a finite zero-sum two person $G(X; \mathscr{S})$, called "hide and seek, $X$;$\mathscr{S}$." Player A "hides" in $X$ by choosing a point $x \in X$; player $B$ "seeks" player $A$ by selecting one of the sets $S_j$. Player $A$ must pay $B$ one unit if $x \in S_j$; otherwise the payoff is zero. Now if $X$ is finite, $G(X; \mathscr{S})$ has a value $v(X; \mathscr{S})$ which can be described as follows. If $Y$ is a finite set, let $\mathscr{A}(Y)$ be the class of all probability distributions on $Y$. The fundamental theorem of game theory (David (1960), Theorem 6.6) implies

$$v(X;\ \ ) = \max_{Q\in\mathscr{A}(\mathscr{S})} \min_{x\in X} Q(\mathrm{Star}(x)) = \min_{P\in\mathscr{A}(X)} \max_{S\in\mathscr{S}} P(S),$$

where $\mathrm{Star}\,(x) = \{S \in \mathscr{S} \mid x \in S\}$.

This means that $B$ can win at least $v(X; \mathscr{S})$, on the average, no matter what strategy $A$ uses, provided he selects his sets $S$ according to the probability distribution $Q$ which achieves $v$. Conversely $A$ can assure himself of losing no more

than $v(X; \mathscr{S})$, on the average, by selecting his point $x$ according to a distribution $P$ which achieves $v$. The distributions $Q$ and $P$ are called *optimal (mixed) strategies.*

In Section 2 we shall show the surprising result

$$I_{\mathscr{S}}(X) = \log\left(v(X; \mathscr{S})\right)^{-1} = \max_{P \in \mathscr{A}(X)} I_{\mathscr{S};P}(X),$$

which is the main purpose of the paper. This determines $I_{\mathscr{S}}(X)$ constructively, and shows that, in the worst case when "Nature" has chosen that $P$ which maximizes $I_{\mathscr{S};P}(X)$ for $P$ in $\mathscr{A}(X)$, no bits can be saved in storing outcomes of $X$ by taking into account the distribution $P$ of outcomes and using optimum variable-length coding with minimum expected word length. For $I_{\mathscr{S};P}(X)$ is the number of bits necessary to describe $X$ acceptably with arbitrarily long codewords when the outcomes are distributed according to $P$ and $P$ is taken into account in coding the outcomes; whereas $I_{\mathscr{S}}(X)$ is the number of bits necessary if probability is ignored. Thus for the $P$ achieving the above inequality, nothing can be saved if arbitrarily long code words can be used. (However, if words of bounded length only can be used, arbitrarily large savings can indeed accrue, as we shall see in the closing section.)

## 2. The main theorem for finite sets.

THEOREM 1. *If $X$ is finite,*

$$I_{\mathscr{S}}(X) = \log\left(v(X; \mathscr{S})\right)^{-1} = \max_{P \in \mathscr{A}(X)} I_{\mathscr{S};P}(X).$$

*The distribution $P$ achieves this maximum if and only if it is an optimum mixed strategy for the hider in $G(X; \mathscr{S})$. In addition, any optimum mixed strategy $Q$ for the seeker in $G(X; \mathscr{S})$ can be used to select a "random code" which comes arbitrarily close to achieving $I_{\mathscr{S}}(X)$ for large enough $N$.*

REMARK. We note that it is possible that

$$H_{\mathscr{S}}(X) > \max_{P \in \mathscr{A}(X)} H_{\mathscr{S};P}(X).$$

For example, if $X$ is the set of vertices of a pentagon, and if $\mathscr{S}$ consists of the pairs of adjacent vertices of $X$, it is obvious that

$$H_{\mathscr{S}}(X) = \log 3$$

but it is not hard to show that

$$\max_P H_{\mathscr{S};P}(X) = \log 5 - \tfrac{4}{5}\log 2,$$

which is achieved when the probability of each vertex is $\frac{1}{5}$.

PROOF OF THE THEOREM. Clearly $I_{\mathscr{S}}(X) \geqq I_{\mathscr{S};P}(X)$ for any $P$, since the (Shannon) entropy of $M$ probabilities is always less than or equal to $\log M$. Also, since

$$\log M \geqq \sum_{i=1}^{\infty} q_i \log {q_i}^{-1} \geqq \log \frac{1}{\max_i q_i}$$

it follows that

$$I_{\mathscr{S};P}(X) \geqq \log \frac{1}{\max_{S \in \mathscr{S}} P(S)},$$

and so

$$I_{\mathscr{S}}(X) \geqq \log(\min_P \max_S P(S))^{-1} = \log(v(X;\mathscr{S}))^{-1}.$$

To get the other inequality, we shall first show that

$$I_{\mathscr{S}}(X) \leqq \log(\min_x Q(\mathrm{Star}(x))^{-1} = \log q^{-1},$$

say, for any $Q$ in $\mathscr{A}(\mathscr{S})$.

For a fixed $N$ and $M$, independently choose $M$ sets $\mathbf{S}_1, \cdots, \mathbf{S}_M$ from $\mathscr{S}^N$ according to the distribution $Q^N$. For a fixed $\mathbf{x} \in x^N$, we have

$$\Pr(\mathbf{x} \text{ covered}) = Q^N(\mathrm{Star}(x)) \geqq q^N, \quad \text{i.e.,}$$

$$Q^N(\mathbf{x} \notin \mathbf{S}_i) \leqq 1 - q^N.$$

By independence,

$$Q^N(\mathbf{x} \notin \bigcup_{i=1}^M \mathbf{S}_i) \leqq (1-q^N)^M.$$

Thus, since there are $n^N$ possible $\mathbf{x}$'s,

$$Q^N\{\text{for some } \mathbf{x}, \mathbf{x} \notin \bigcup_{i=1}^M S_i\} \leqq n^N(1-q^N)^M.$$

Hence if $n^N(1-q^N)^M < 1$, there will be at least one choice of the $\mathbf{S}_i$ for which $X^N = \cup\, \mathbf{S}_i$. But if $M > -N \log n / \log(1-q^N)$, this will be the case. Thus

$$I_{\mathscr{S}}(X) \leqq \frac{1}{N} \log\left(\frac{-N \log n}{\log(1-q^N)} + 1\right) \leqq \frac{1}{N} \log\left(\frac{N \log n}{q^N(1+O(q^N))} + 1\right)$$

$$\leqq \frac{1}{N}\left\{\log q^{-N} + \log\left(\frac{N \log n}{1+O(q^N)} + q^N\right)\right\}.$$

Letting $N \to \infty$, we obtain

$$I_{\mathscr{S}}(X) \leqq \log q^{-1},$$

as promised.

If we now minimize this bound over $Q$, we obtain

$$I_{\mathscr{S}}(X) \leqq \log(\max_Q \min_x Q(\mathrm{Star}(x))^{-1},$$

which also equals $\log 1/v(X;\mathscr{S})$, so that equality follows.

To complete the proof, it remains to show $I_{\mathscr{S}}(X) = \max_P I_{\mathscr{S};P}(X)$. However, this follows immediately from the inequality

$$I_{\mathscr{S}}(X) \geqq I_{\mathscr{S};P}(X)$$

mentioned above, and the inequality

$$I_{\mathscr{S};P}(X) \geqq \log(\max_S P(S))^{-1}$$

when $P$ is chosen to minimize $\max_S P(S)$.

COROLLARY. *Let* $X, \mathscr{S}$; $Y, \mathscr{T}$, *be as in Theorem* 1, *and let* $\mathscr{S} \times \mathscr{T}$ *denote the collection of subsets of* $X \times Y$ *of the form* $S \times T$, $S \in \mathscr{S}$, $T \in \mathscr{T}$. *Then*

$$I_{\mathscr{S} \times \mathscr{T}}(X \times Y) = I_{\mathscr{S}}(X) + I_{\mathscr{T}}(Y).$$

PROOF. By Theorem 1, we need only prove

$$v(X \times Y; \mathscr{S} \times \mathscr{T}) = v(X; \mathscr{S}) v(Y; \mathscr{T}).$$

To prove this, use for the seeker the strategy $Q_X \times Q_Y$, where $Q_X \times Q_Y$ denotes product measure, an element of $\mathscr{A}(\mathscr{S} \times \mathscr{T})$, with $Q_X$ his optimal strategy in $X; \mathscr{S}$, and $Q_Y$ his optimal strategy for $Y; \mathscr{T}$. This proves that

$$v(X \times Y; \mathscr{S} \times \mathscr{T}) \geqq v(X; \mathscr{S}) v(Y; \mathscr{T}).$$

Conversely, use for the hider $P_X \times P_Y$, an element of $\mathscr{A}(X \times Y)$, where $P_X$ is optimal for $X; \mathscr{S}$ and $P_Y$ for $Y; \mathscr{T}$. This shows

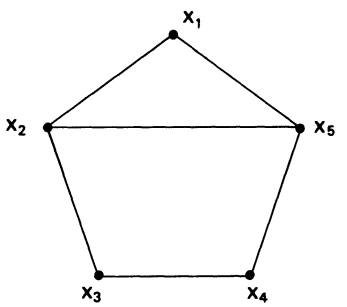
$$v(X \times Y; \mathscr{S} \times \mathscr{T}) \leqq v(X; \mathscr{S}) v(Y; \mathscr{T}),$$

and so proves the corollary.

*Application to graph theory.* Theorem 1 allows us to calculate the absolute entropy of a finite undirected graph, relative either to the class of cliques or to the class of talons. Cliques are sets of diameter 1, and talons are spheres of radius 1. Here the metric is the one in which the distance between two vertices is the length of the shortest path connecting them, or one more than the number of vertices if there is not a path connecting them. Thus the distance between adjacent vertices is 1.

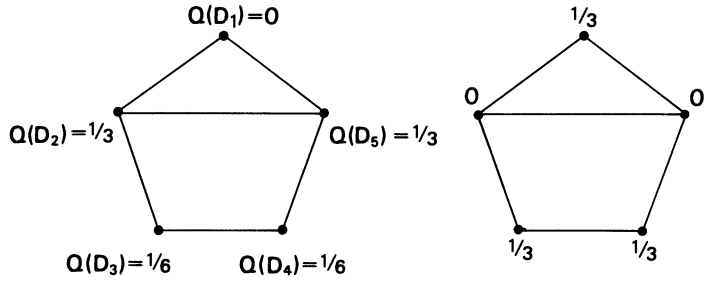
Our corollary shows that the function I has the important property of being additive under the "Kronecker product" of graphs. Thus, the absolute entropy of a graph can be used to decide whether a given graph is the Kronecker product of two given graphs: if the absolute entropies can be calculated rapidly, in some cases the Kronecker product property can be immediately ruled out.

For an example of the absolute entropies of a graph, consider the graph below:




The maximal cliques are $C_1 = \{x_1, x_2, x_5\}$, $C_2 = \{x_2, x_3\}$, $C_3 = \{x_3, x_4\}$, and $C_4 = \{x_4, x_5\}$. If $\mathscr{S}$ is the set of (maximal) cliques, $I_{\mathscr{S}}(X) = \log 2$, since the probability assignment $p_2 = p_3 = p_4 = p_5 = \frac{1}{4}$, $p_1 = 0$ makes $P(S) = \frac{1}{2}$, all

$S \in \mathscr{S}$. And, letting $C_1$ and $C_2$ have probability $\frac{1}{2}$ gives a $Q$ with $Q(\text{Star}(x)) = \frac{1}{2}$, all $x$. The talons are $D_1 = \{x_1, x_2, x_5\}$, $D_2 = \{x_2, x_1, x_5, x_3\}$, $D_3 = \{x_3, x_2, x_4\}$, $D_4 = \{x_4, x_3, x_5\}$, $D_5 = \{x_5, x_1, x_4, x_2\}$. Here $I_{\mathscr{S}}(X) = \log \frac{3}{2}$, because of the following pair of probability assignments:




MIN Q = 2/3 MAX P = 2/3

We close this section with some remarks on the connection between absolute entropy and integer linear programming. In Balinsky (1968), page 214, an integer linear programming problem is discussed, which arises from a minimum covering problem. Namely, let $A$ be the $n \times m$ zero-one matrix whose $(i, j)$th entry is 1 if and only if the $i$th element of $X$ is in $S_j$; otherwise the entry is zero. Let $e_k$ be the vector of $k$ ones. The integer program is

(P) minimize $(\xi, e_m)$, $\xi$ an $m$-vector of nonnegative integers, subject to

$$A\xi \geqq e_n$$

($\geqq$ for vectors means that the inequality holds componentwise). Observe that the minimum exists, and $\xi_j = 0$ or 1, $1 \leqq j \leqq m$, at minimum $\xi$.

For $\xi$ an $m$-vector, let $S(\xi)$ be the subset of $\mathscr{S}$ containing those $S_j$ for which $\xi_j$ is 1. For $\mathscr{S}_1 \subset \mathscr{S}$, let $\chi(\mathscr{S}_1)$ be the $m$-vector whose $j$th component is 1 if $S_j \in \mathscr{S}_1$, and 0 otherwise. Then

$$A\xi \geqq e_n \Leftrightarrow S(\xi) \text{ covers } X;$$

$$\mathscr{S}_1 \text{ covers } X \Leftrightarrow A\chi(\mathscr{S}_1) \geqq e_n.$$

Thus, the integer program $P$ is equivalent to the problem of covering $X$ with a subset of $\mathscr{S}$ of smallest cardinality; minimizing $\xi$ leads to such a minimum cover, and conversely.

Now (Gale (1960), Section 7.1) the Fundamental Theorem of the Theory of Games can be proved from Linear Programming Duality in the following way, with $A$ the same as above; $A$ is also the payoff matrix in hide and seek, $(X; \mathscr{S})$.

Let $F$ be the linear programming problem (not in integers)

(F) minimize $(\xi, e_m)$ in $\xi \geqq 0$ subject to $A\xi \geqq e_n$;

its dual is

(G) maximize $(\eta, e_m)$ in $\eta \geqq 0$ subject to $\eta A \leqq e_m$.

From the fact (Gale (1960) Chapter 3) that the minimum for $F$ is the maximum for $G$ (the *value* of the program, clearly positive), it is proved that the game has a value, the reciprocal of the value of the program. Furthermore, any minimizing $\xi$ normalized to have sum of entries 1 by dividing by the value of the program is an optimal strategy for the hider. Likewise, any $\eta$ so normalized is optimal for the seeker, and every pair of optimal strategies arises in this way. (In fact, this is usually the way we would find the value and strategies for complicated $A$.)

The whole point of this discussion is that the integer program $P$ and the program $F$ differ only in that in $P$, $\xi$ is required to have integral components. And $P$ was the program for determining the one-shot $\mathscr{S}$-entropy of $X$, whereas $F$ determined the absolute $\mathscr{S}$-entropy. Rarely does the unrestricted version of an integer program have such a nice interpretation related so neatly to the interpretation of the original problem.

The program $F$ can be thought of as the problem of extracting a minimum subcover of $\mathscr{S}$, where each $S_j$ in $\mathscr{S}$ can be counted fractionally, but each $x$ in $X$ is to be covered, in the sense that the sum of the fractional coefficients given to each of the $S_j$ containing the given $x$ is at least 1. The fact that the log of this minimum is also the absolute entropy is the content of Theorem 1.

**3. Compact metric spaces.** In this section, we shall briefly consider the problem of calculating $I_{\mathscr{S}}(X)$ where $X$ is an arbitrary compact metric space, and $\mathscr{S}$ is either the class of closed subsets of $X$ of radius at most $\varepsilon/2$, or the class of closed subsets of diameter at most $\varepsilon$. Our results are incomplete, in that we shall prove the analogue of Theorem 1 only for all but a countable number of $\varepsilon$ (depending on $X$).

Thus, let $R(\varepsilon)$ be the class of closed subsets of $X$ which have radius at most $\varepsilon/2$ (the radius of a set is the radius of the smallest sphere containing the set), and $D(\varepsilon)$ be those closed sets of diameter $\leqq \varepsilon$.

As in the finite case, we can define the two games "hide and seek $(X; \mathscr{S})$" where $\mathscr{S}$ = either $R(\varepsilon)$ or $D(\varepsilon)$. However, to define the set $\mathscr{A}(\mathscr{S})$ of probability distributions on the set $\mathscr{S}$ it is convenient to make $\mathscr{S}$ itself into a compact metric space by means of the *Hausdorff metric* (Hausdorff (1957), page 28). Thus if $F$, $G \in \mathscr{S}$, define

$$d_H(F, G) = \max\left(\max_{x \in F} \min_{y \in G} d(x, y), \max_{y \in G} \min_{x \in F} d(x, y)\right).$$

It is shown in Hausdorff (1957) that under this metric $\mathscr{S}$ does indeed become a compact metric space. And it can also be shown (McEliece and Posner (1971)) that the game $G(X; \mathscr{S})$ does have a value $v(X; \mathscr{S})$ which is given by

$$(*) \qquad v(X; \mathscr{S}) = \inf_{\mu \in \mathscr{A}(X)} \max_{F \in \mathscr{S}} \mu(F) = \max_{\mu \in \mathscr{A}(\mathscr{S})} \inf_{x \in X} \bar{\mu}(\mathrm{Star}(x)).$$

If $\mathscr{S} = R(\varepsilon)$ or $D(\varepsilon)$, we abbreviate $v(X; \mathscr{S})$ as $v(\varepsilon)$, call $\lim_{\mu \downarrow 0} v(\varepsilon - \eta) = v(\varepsilon -)$, and write $I_{\mathscr{S}}(X)$ as $I_\varepsilon(X)$ or even as $I_\varepsilon$.

REMARK. One may wish to define the situation for $R(\varepsilon)$ by measures on $X$, where the payoff to the seeker is 1 if and only if the distance between the hider's chosen

point and that of the seeker is at most $\varepsilon/2$. However, that the two approaches are equivalent follows from a known selection theorem (Parthasarathy (1967)).

THEOREM 2. *For all* $\varepsilon > 0$,

$$\log v(\varepsilon)^{-1} \leqq I_\varepsilon \leqq \log v(\varepsilon-)^{-1}.$$

*Consequently, with at most countable many exceptions,* $I_\varepsilon = \log 1/v(\varepsilon)$. *In any case, for every* $\varepsilon$ *we have* $\log 1/v(\varepsilon) = \sup_{P \in \mathscr{M}(X)} I_{\mathscr{S};P}(X)$, *and* $v(\varepsilon)$ *is continuous from above in* $\varepsilon$.

PROOF. We treat only the case $\mathscr{S} = R(\varepsilon)$, the case $\mathscr{S} = D(\varepsilon)$ being entirely analogous.

First of all we note that the inequality

$$\log v(\varepsilon)^{-1} \leqq I_\varepsilon \tag{1}$$

remains true in the infinite case; the proof is the same as in Theorem 1.

The given $\varepsilon$ is fixed. Next, fix $\eta$ with $\varepsilon - \eta > 0$. Let $J$ be a partition of $X$ into a finite number of subsets, each of which can be enclosed in a sphere of radius $\eta/2$. Let $\mathscr{S}(J)$ be the collection of subsets of $X$ which are unions of $J$-sets and are also in $R(\varepsilon) = \mathscr{S}$. Then $\mathscr{S}(J)$ can be regarded as a collection of subsets of the finite set $J$. Denote the absolute entropy of this finite space by $I_{\mathscr{S}(J)}$, and the value of the corresponding game by $v(J)$. Then

$$I_\varepsilon \leqq I_{\mathscr{S}(J)}, \tag{2}$$

since the sets in $\mathscr{S}(J)$ are also in $\mathscr{S}$.

Next we claim

$$v(J) \geqq v(\varepsilon-\eta). \tag{3}$$

To see this, let $P$ be a probability distribution on $J$ which achieves $v(J) = \min_P \max_{S \in \mathscr{S}(J)} P(S)$. We extend $P$ to a probability on $X$ by choosing a point $s$ in each set $j$ of the partition $J$ which is the center of a sphere of radius $\varepsilon/2$ containing $j$, and assigning this point $s$ the probability $P(j)$; the desired measure is the atomic one concentrated at these points. If $Y$ is any set in $R(\varepsilon-\eta)$, then the union of those $J$-sets containing the points of positive probability in $Y$ will be a set in $\mathscr{S}(J)$. Consequently $\max_{S \in \mathscr{S}(J)} P(S) \geqq \sup_{Y \in R(\varepsilon-\eta)} P(Y)$, and so

$$\begin{aligned} v(J) &= \min_{P \in \mathscr{M}(J)} \max_{S \in \mathscr{S}(J)} P(S) \geqq \inf_{P \in \mathscr{M}(X)} \sup_{Y \in R(\varepsilon-\eta)} P(Y) \\ &= v(\varepsilon-\eta), \end{aligned}$$

since the right-hand "inf" is over a larger set of probabilities. Combining (1), (2), and (3), we find

$$\log v(\varepsilon)^{-1} \leqq I_\varepsilon \leqq I_{\mathscr{S}(J)} = \log v(J)^{-1} \leqq \log v(\varepsilon-\eta)^{-1}. \tag{4}$$

Let $\eta \downarrow 0$ to obtain the inequality $I_\varepsilon \leqq \log 1/v(\varepsilon-)$.

To prove the last sentence of Theorem 2, note that as in the proof of Theorem 1

$$I_{\mathscr{S};P}(X) \geqq \log \frac{1}{\sup_S P(S)}$$

for every $P \in \mathscr{A}(X)$, so that

$$\sup_P I_{\mathscr{S};P}(X) \geqq \log \frac{1}{v(\varepsilon)} = \sup_P \log \frac{1}{\sup P(S)}.$$

Let us abbreviate this as $r(\varepsilon) \geqq s(\varepsilon)$; both $r$ and $s$ are decreasing functions of $\varepsilon$. Then (4) implies

$$s(\varepsilon) \leqq r(\varepsilon) \leqq I_\varepsilon \leqq s(\varepsilon-).$$

We shall prove $r(\varepsilon) = s(\varepsilon)$ by showing that $r(\varepsilon)$ is continuous from above in $\varepsilon$. Then if $\delta_n$ is any sequence of positive reals which decreases to zero such that for all $n$, $r(\varepsilon+\delta_n) = s(\varepsilon+\delta_n)$, then as $n \to \infty$ we see that $r(\varepsilon) = s(\varepsilon+) \leqq s(\varepsilon)$, which implies $r(\varepsilon) = s(\varepsilon)$.

Thus it remains to show that $r(\varepsilon) = \sup_P I_{\mathscr{S}(\varepsilon);P}(X)$ is continuous from above in $\varepsilon$. First of all, it is known (Posner and Rodemich (1971), Parthasarathy (1967), Corollary 1) that for fixed $P$, $I_{\mathscr{S};P}$ is continuous from above in $\varepsilon$. Now for a fixed $\eta > 0$ pick $P$ such that

$$I_{\varepsilon;P} > \sup I_{\varepsilon;P} - \eta/2 \tag{5}$$

where $I_{\varepsilon;P} = I_{\mathscr{S}(\varepsilon);P}$. Next pick $\delta$ such that for $\delta' \leqq \delta$

$$I_{\varepsilon+\delta';P} > I_{\varepsilon;P} - \eta/2, \tag{6}$$

using the continuity of $I_{\varepsilon;P}$ from above. Combining (5) and (6) we have

$$I_{\varepsilon+\delta';P} > \sup_P I_{\varepsilon;P} - \eta$$

for all $\delta' \leqq \delta$. *A fortiori*, for $\delta'$ sufficiently small depending on $\eta$ we have

$$\sup_P I_{\varepsilon+\delta';P} > \sup_P I_{\varepsilon;P} - \eta,$$

which proves the continuity of $\sup_P I_{\varepsilon;P}$ from above. This completes the proof.

Note that Theorem 2 implies $I = \log 1/v(\varepsilon)$ if and only if $I_\eta$ is continuous from above in $\eta$ at $\eta = \varepsilon$. This is a special case of the open conjecture we discuss in the next section, which implies the result for metric subspaces of a Hilbert Space.

## 4. Concluding remarks.

1. We conjecture that $I_\varepsilon = \log 1/v(\varepsilon)$, and point out that there is no hope of proving this in general by showing that $v(\varepsilon)$ is continuous from below, since $v(\varepsilon)$ frequently fails to be continuous from below. However, $v(\varepsilon)$ is continuous in $\varepsilon$ for many interesting choices of $X$: if $X$ is a compact homogeneous space

under an invariant metric and the invariant probability, the invariant probability of a sphere of radius $\varepsilon/2$ is continuous in $\varepsilon$, and that probability measure is easily seen to be an optimal strategy. Here $v(\varepsilon) = \Pr$ (sphere radius $\varepsilon/2$). Thus $n$-spheres in Eulidean space, Lie groups, etc., are all included here. We can also prove our conjecture if $X$ can be embedded isometrically a normed linear space with the property that every point on the boundary of the unit ball is an extreme point of the unit ball (e.g., a Hilbert space). We omit the proof because it seems to shed no light on the general case. (Such spaces are called *strictly convex*.)

2. We can now give an example of a sequence of $X_n$ such that

$$I_\varepsilon(X_n) \to \log 2$$

but

$$H_\varepsilon(X_n) \to \infty.$$

Furthermore, the example has the property that

$$I_\varepsilon(X_n) = \max_P I_{\varepsilon;P}(X_n).$$

Thus, although Nature can prevent us from taking advantage of knowledge of $P$ if arbitrarily long blocks of data are allowed to be encoded, we can still take advantage of $P$ if no storage is allowed, that is, if we are forced to use one-shot entropies. For by Posner and Rodemich, Theorem 2,

$$H_{\varepsilon;P}(X) \leqq I_{\varepsilon;P}(X) + \log^+ I_{\varepsilon;P}(X) + C,$$

$C$ a universal constant (this result is crucial in proving the continuity from above in $\varepsilon$ of $I_{\varepsilon;P}(X)$, used in Section 3). Thus, $H_{\varepsilon;P}(X_n)$ is bounded, whereas $H_\varepsilon(X_n) \to \infty$. The example can be modified using techniques of Section 3 to make each $X_n$ a finite set.

Here is the example. Let $X_n$ be an $n$-sphere, with geodesic metric and invariant probability. Let $\varepsilon$ be such that the sphere of radius $\varepsilon/2$ has surface area $(\frac{1}{2}) - \delta_n$ say, where $0 \leqq \delta_n \leqq \frac{1}{2}$ and $\delta_n \to 0$; the surface area of $X_n$ is normalized to 1.

An easy lemma gives

$$H_\varepsilon(X_n) \geqq \log(n+2)$$

(also, $H_\varepsilon(X_n) \leqq \log 2n$). Furthermore, the above remarks on homogeneous spaces give

$$I_\varepsilon(X_n) = \log(\tfrac{1}{2} - \delta_n)^{-1},$$

and so

$$I_\varepsilon(X_n) \to \log 2,$$

as promised.

**Note added in proof.** The conjecture in 1 above is false, but true for subspaces of normed linear spaces.

## REFERENCES

LORENTZ, G. G. (1966). *Approximation of Functions.* Holt, Rinehart and Winston, New York.

POSNER, E. C. and RODEMICH, E. R. (1971). Epsilon entropy and data compression. Submitted to *Ann. Math. Statist.*

BALINSKY, M. L. (1968). Inter programming: Methods, uses, computation. *Mathematics of the Decision Sciences, Part* 1. (George B. Dantzig and Arthur F. Veinott, Jr., eds.) 179–256. Vol. 11 in Lectures in Applied Mathematics, Amer. Math. Soc.

BERGE, CLAUDE (1962). *The Theory Graphs and its Applications* (Translation). Wiley, London.

GALE, DAVID (1960). *The Theory of Linear Economic Models.* McGraw-Hill, New York.

HAUSDORFF, F. (1957). *Set Theory.* Chelsea, New York. (Reprinted.)

MCELIECE, R. J. and POSNER, E. C. (1971). Hiding in a compact mètric space. In preparation.

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

PROKHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. (English translation in *Theory of Probability and its Applications* **1** 157–214.)