Journal of Statistical Physics, Vol. 1, No. 1, 1969

Differential Entropy and Tiling

Edward C. Posner[1] and Eugene R. Rodemich[1]

Received February 17, 1969

This paper relates the differential entropy of a sufficiently nice probability density function p on Euclidean n-space to the problem of tiling n-space by the translates of a given compact symmetric convex set S with nonempty interior. The relationship occurs via the concept of the epsilon entropy of n-space under the norm induced by S, with probability induced by p. An expression is obtained for this entropy as ϵ approaches 0, which equals the differential entropy of p, plus n times the logarithm of $2/\epsilon$, plus the logarithm of the reciprocal of the volume of S, plus a constant $C(S)$ depending only on S, plus a term approaching zero with ϵ. The constant $C(S)$ is called the entropic packing constant of S; the main results of the paper concern this constant. It is shown that $C(S)$ is between 0 and 1; furthermore, $C(S)$ is zero if and only if translates of S tile all of n-space.

KEY WORDS: Differential entropy; Tiling; Entropy; Close packing; Random coding; Convex sets; Epsilon entropy; Information-theoretic geometry.

1. INTRODUCTION

This paper defines a constant $C(S)$ for a compact convex symmetric set S in Euclidian n-space E^n having nonempty interior, such that

$$0 \leqslant C(S) \leqslant 1$$

such that $C(S)$ is a continuous function of S in a natural topology on the space of S, and such that

$$C(S) = 0$$

if and only if translates of S tile E^n, that is, if and only if E^n can be covered by a union of translates of S such that the intersection of any two such translates does not contain

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.

[1] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California.

an open set. This constant $C(S)$ is called the *entropic packing* of S, and is defined via the notion of the *differential entropy* of a sufficiently nice density function on E^n. The differential entropy occurs as one term in an expression for the *epsilon entropy* of E^n under the probability distribution induced by p, with metric defined by the norm induced by S.

We will now define these terms. First we define the notion of the epsilon entropy $H_\epsilon(X)$ of a complete separable metric space X under a probability measure μ such that the open sets of X are measurable; that is, we define the epsilon entropy of a *probabilistic metric space.*[1,2] The entropy is defined as the infimum (actually minimum) of the entropies of all partitions of X by sets of diameter at most ϵ. The entropy of a partition is defined as

$$\sum p_i \log(1/p_i) \tag{1}$$

where

$$p_i = \mu(U_i)$$

the probability of the ith set of the partition.

Now let p be a density function on E^n. Then the differential entropy $H(p)$ of the density p is defined as[3]

$$\int p(x) \log[1/p(x)]\, dm(x)$$

where $dm(x)$ is Lebesgue measure on E^n. This integral is either finite or $-\infty$, since

$$p \log (1/p) \leqslant 1/e$$

Finally, if S is a compact convex symmetric set in E^n with nonempty interior, such that the origin O is its center of symmetry, then the norm

$$\| \cdots \|_S$$

on E^n is defined as

$$\| x \|_S = \min\{\lambda > 0 / x \in \lambda S\}$$

where λS is the set of all λs, $s \in S$. Then E^n is a complete normed linear space under $\| \cdots \|_S$, and this norm is equivalent to the Euclidean one.

We are interested in the probabilistic metric space X whose point set is E^n, whose metric is induced by $\| \cdots \|_S$, and whose probability measure $d\mu$ is defined by

$$d\mu = p(x)\, dm(x)$$

and $p \geqslant 0$, $\int p dm = 1$.

We want an expression for $H_\epsilon(X)$ valid as $\epsilon \to 0$. To do this, we need to assume that p is nice in a sense to be made precise. Let v_1 be the Euclidean volume of S. Then we prove that for all S and for a certain class of p not depending on S, we have

$$H_\epsilon(X) = n \log(2/\epsilon) + H(p) + \log(1/v_1) + C(S) + o(1) \tag{2}$$

as $\epsilon \to 0$, where $o(1)$ may depend on p and S. The constant $C(S)$ occurring in this equation is the entropic packing constant of S; it satisfies, as we shall show,

$$0 \leqslant C(S) \leqslant 1$$

and furthermore $C(S) = 0$ if and only if translates of S tile E^n. We also show that $C(S)$ is a continuous function of S in a natural sense. The proof that $C(S) \leqslant 1$ relies heavily on a random coding argument in Reference 2; except for this, our discussion here is reasonably self-contained.

2. THE n-CUBE UNDER LEBESGUE MEASURE

This section introduces the entropic packing constant $C(S)$ in terms of partitions of the n-cube by measurable sets of diameter at most ϵ under $\| \cdots \|_S$.

We need one more definition, that of the *Hausdorff metric* on the space of compact subsets of a given complete separable metric space (pp. 166–172 of Reference 4). For two closed sets A and B, define the Hausdorff distance ρ between A and B as

$$\rho(A, B) = \max\{\sup_{x\in A} \inf_{y\in B} d(x, y), \sup_{y\in B} \inf_{x\in A} d(x, y)\} \tag{3}$$

The space R of compact subsets of the complete separable metric space X with metric d is then itself a complete separable metric space with metric ρ, and R is compact under ρ if X is compact under d. We then have Theorem 1.

Theorem 1. Let $\| \cdots \|_S$ be the norm on E^n associated with a compact convex symmetric set in E^n with nonempty interior. Let v_1 be the (Lebesgue) volume of S. If $\mathscr{U} = \{U_j\}$ is any partition of E^n into sets of diameters $\leqslant\epsilon$ (under $\| \cdots \|_S$), and J is the set of integers j for which U_j lies wholly in the cube

$$0 < x_k < L, \qquad k = 1,\ldots, n \tag{4}$$

where x_k is the kth coordinate in E^n, then

$$\sum_{j\in J} m(U_j) \log \frac{1}{m(U_j)} \geqslant L^n \left[\log \frac{2^n}{\epsilon^n v_1} + C(S) + g_1\left(\frac{\epsilon}{L}\right)\right] \tag{5}$$

where $C(S)$ is a constant depending only on S, and $g_1(t)$ is a function depending only on S, such that $g_1(t) \to 0$ as $t \to 0^+$. Also, there is an ϵ-partition $\mathscr{V} = \{V_j\}$ of the cube (4) with

$$\sum_{j} m(V_j) \log \frac{1}{m(V_j)} = L^n \left[\log \frac{2^n}{\epsilon^n v_1} + C(S) + g_2\left(\frac{\epsilon}{L}\right)\right] \tag{6}$$

where $g_2(t)$ is a function depending only on the metric, with $g_2(t) \to 0$ as $t \to 0^+$.

The constant $C(S)$, the *entropic packing constant*, is at most 1, is nonnegative, and is zero if and only if translates of S tile E^n. Furthermore, $C(S)$ is continuous in the Hausdorff metric on the space of compact subsets of E^n.

Proof. We shall need the following known result (Reference 5, Sec. 47), a consequence of the Brün–Minkowski lemma:

Let S be a compact convex set in Euclidean n-space with nonempty interior. Then the volume of S is at most the volume of $[S + (-S)]/2$, the symmetric convex set consisting of all $(x - y)/2$, $x, y \in S$. Equality holds if and only if S itself is symmetric, i.e., if and only if $S = -S$.

As a corollary to this result, we note that the given convex symmetric set S has Euclidean volume equal to or greater than that of any set B in the space of diameter $\leqslant 2$, with equality if and only if B differs from S by a translation. To show this, we can assume that B is symmetric, since $[B + (-B)]/2$ is symmetric, has diameter at most 2, and at least as much volume as B. Now

$$\| x \|_S \leqslant 1 \Leftrightarrow x \in S$$

Hence, if B is symmetric, has diameter 2, and $x, y \in B$, then

$$\begin{aligned} \| x - y \|_S &\leqslant 2 \\ \| x/2 - y/2 \|_S &\leqslant 1 \\ x/2 - y/2 &\in S \\ [B + (-B)]/2 &\subset S \\ B &\subset S \end{aligned}$$

as required.

It then follows that if B is a closed convex set in Euclidean n-space of $\| \cdots \|_S$ diameter at most 1, and if B has volume close to the volume of S, then B itself is close to S. That is, the volume of the symmetric difference between some translate of B and S must be small. This result follows from the fact that the space of closed sets contained in some fixed sphere of Euclidean n-space is compact under the Hausdorff metric. This fact will be used later to prove that $C(S)$ is continuous in the Hausdorff metric.

Let X denote the probabilistic metric space consisting of the unit cube

$$0 < x_k < 1, \qquad 1 \leqslant k \leqslant n$$

in E^n, under the metric induced by $\| \cdots \|_S$, and Lebesgue measure as its probability distribution. We wish to consider

$$D(\epsilon) = H_\epsilon(X) - \log(2^n/\epsilon^n v_1) \tag{7}$$

Note that

$$D(\epsilon) \geqslant 0, \qquad \text{all } \epsilon \tag{8}$$

since, as we have seen, the maximum probability of an ϵ-set in X is the probability of the set $\frac{1}{2}\epsilon S$.

We claim that $D(\epsilon)$ approaches a finite limit as $\epsilon \to 0$. First, for ϵ sufficiently large, the diameter of X is less than ϵ, and $H_\epsilon(X) = 0$. Hence for large ϵ, $D(\epsilon) < \infty$.

Now let q be a positive integer. We cut the n-cube into q^n equal cubes of side $1/q$. The small cubes with a uniform measure form probabilistic metric spaces $X_{(l)}$, $1 \leqslant l \leqslant q^n$. Then (Reference 2, Sec. 5)

$$H_{\epsilon/q}(X) \leqslant \sum_{l=1}^{q^n} \frac{1}{q^n} H_{\epsilon/q}(X_{(l)}) + \log(q^n)$$

$$= H_\epsilon(X) + \log(q^n)$$

$$H_{\epsilon/q}(X) - \log(2^n q^n/\epsilon^n v_1) \leqslant H_\epsilon(X) - \log(2^n/\epsilon^n v_1)$$

or

$$D(\epsilon/q) \leqslant D(\epsilon) \tag{9}$$

This inequality shows first that $D(\epsilon) < \infty$ for all $\epsilon > 0$. Now for $\delta > 0$, choose ϵ_1 so that

$$D(\epsilon_1) \leqslant \liminf_{\epsilon \to 0} D(\epsilon) + \delta \tag{10}$$

For q a positive integer and $\epsilon_1/q \geqslant \epsilon > \epsilon_1/(q+1)$,

$$\begin{aligned} D(\epsilon) &= H_\epsilon(X) - \log(2^n/\epsilon^n v_1) \\ &\leqslant H_{\epsilon_1/(q+1)}(X) - \log(2^n q^n/\epsilon_1{}^n v_1) \\ &= D(\epsilon_1/(q+1)) + n \log[(q+1)/q] \end{aligned}$$

By (9), we have

$$D(\epsilon) \leqslant D(\epsilon_1) + n \log[(q+1)/q], \qquad \epsilon_1/q \geqslant \epsilon > \epsilon_1/(q+1)$$

Hence, by (10),

$$\limsup_{\epsilon \to 0} D(\epsilon) \leqslant D(\epsilon_1) \leqslant \liminf_{\epsilon \to 0} D(\epsilon) + \delta$$

Letting $\delta \to 0$, we see that $D(\epsilon)$ has a limit as $\epsilon \to 0$.

Let

$$C(S) = \lim_{\epsilon \to 0} D(\epsilon) \tag{11}$$

and define

$$g_2(\epsilon) = D(\epsilon) - C(S)$$

Then $g_2(0^+) = 0$, and Eq. (6) follows by definition for $L = 1$, if $\{V_j\}$ is an ϵ-partition of entropy $H_\epsilon(X)$.

To show (6) for $L \neq 1$, let $\{W_j\}$ be an ϵ/L partition of the unit cube with entropy $H_{\epsilon/L}(X)$. Then

$$\sum m(W_j) \log[1/m(W_j)] = \log(2^n L^n/\epsilon^n v_1) + C(S) + g_2(\epsilon/L)$$

Take $V_j = LW_j$. Then $\{V_j\}$ is an ϵ-partition of the cube (4). Since

$$m(V_j) = L^n m(W_j)$$

we have

$$\sum m(V_j) \log \frac{1}{m(V_j)} = L^n \left[\sum m(U_j) \log \frac{1}{m(U_j)} + \log \frac{1}{L^n}\right]$$
$$= L^n[\log(2^n/\epsilon^n v_1) + C(S) + g_2(\epsilon/L)],$$

which is (6).

To show (5), it is similarly sufficient to take $L = 1$. Accordingly, let $\{U_j\}$ and J be as stated in the hypotheses, and $L = 1$. Define

$$g_1(\epsilon) = \min \left[0, \inf_u \left\{ \sum_{j \in J} m(U_j) \log \frac{1}{m(U_j)} - \log \frac{2^n}{\epsilon^n v_1} - C(S)\right\}\right] \tag{12}$$

where the inf is taken over all ϵ-partitions U of E^n. Then (5) is satisfied. We only need to show that $g_1(0^+) = 0$.

Suppose that $\epsilon < \frac{1}{2}$. Then all the points of the cube

$$\epsilon < x_k < 1 - \epsilon, \qquad k = 1,..., n \tag{13}$$

have distance at least ϵ from the boundary of the unit cube, and $\{U_j, j \in J\}$ cover (13). Let b be the point with coordinates $(\frac{1}{2}, \frac{1}{2},..., \frac{1}{2})$, and

$$Y_j = \frac{\frac{1}{2}}{\frac{1}{2} - \epsilon} [U_j - b] + b$$

Then $\{Y_j, j \in J\}$ cover the unit cube. Let Z_j be the restriction of Y_j to this cube. As we have seen above, the right side of (6) is the ϵ-entropy of X, for $L = 1$. The sets Z_j have diameters $\leqslant \epsilon/(1 - 2\epsilon)$. Hence

$$\sum_J m(Z_j) \log \frac{1}{m(Z_j)} \geqslant \log \left(\frac{2^n(1 - 2\epsilon)^n}{\epsilon^n v_1}\right) + C(S) + g_2[\epsilon/(1 - 2\epsilon)] \tag{14}$$

Again, for $\epsilon < \epsilon_0$, depending only on S, each of the sets Y_j has measure less than $1/e$. Since the function $t \log(1/t)$ is increasing on the interval $(0, 1/e)$, we then have

$$\sum_J m(Z_j) \log \frac{1}{m(Z_j)} \leqslant \sum_J m(Y_j) \log \frac{1}{m(Y_j)}$$
$$= (1 - 2\epsilon)^{-n} \sum_J m(U_j) \left[\log \frac{1}{m(U_j)} + n \log(1 - 2\epsilon)\right]$$
$$\leqslant (1 - 2\epsilon)^{-n} \left[\sum_J m(U_j) \log \frac{1}{m(U_j)} + n \log(1 - 2\epsilon)\right]$$

Combining this inequality with (14), we obtain

$$\sum_J m(U_j) \log \frac{1}{m(U_j)} - \log \frac{2^n}{\epsilon^n v_1} - C(S)$$
$$\geqslant [(1 - 2\epsilon)^n - 1] \left[\log \frac{2^n}{\epsilon^n v_1} + C(S) + n \log(1 - 2\epsilon)\right]$$
$$+ (1 - 2\epsilon)^n g_1[\epsilon/(1 - 2\epsilon)]$$
$$= g_3(\epsilon)$$

This function $g_3(\epsilon)$ approaches zero as $\epsilon \to 0$. By (12),

$$0 \geqslant g_1(\epsilon) \geqslant \min[0, g_3(\epsilon)]$$

Hence

$$\lim_{\epsilon \to 0} g_1(\epsilon) = 0$$

To prove that $C(S)$ is continuous in the Hausdorff metric, we start with the following observation: Let S be a fixed compact convex set with nonempty interior, symmetric about the origin. Then for any $\delta > 0$, there is an $\alpha > 0$ such that

$$\rho(S, S') < \alpha \Rightarrow (1 + \delta)^{-1} S \subset S' \subset (1 + \delta) S \tag{15}$$

for any convex set S' of this type.

Suppose $\rho(S, S') < \alpha$. Then by (15), any ϵ-partition in the S (or S') metric is a $(1 + \delta) \cdot \epsilon$-partition in the S' (or S) metric. We use again the fact that the right side of (6), for $L = 1$, is $H_\epsilon(X)$. Denote this space in the S' metric by X'. Then, in an obvious notation,

$$H_\epsilon(X) = \log[2^n/\epsilon^n v_1(S)] + C(S) + g_1(\epsilon, S)$$

and since the ϵ-partition which has this entropy is a $(1 + \delta) \cdot \epsilon$-partition of X',

$$H_\epsilon(X) \geqslant H_{\epsilon(1+\delta)}(X') = \log[2^n/\epsilon^n(1 + \delta)^n v_1(S')] + C(S') + g_1[\epsilon(1 + \delta), S']$$

Applying the formula for $H_\epsilon(X)$, we get

$$C(S') - C(S) \leqslant \log\left[\frac{v_1(S')}{v_1(S)}(1 + \delta)^n\right] - g_1[\epsilon(1 + \delta), S'] + g_1(\epsilon, S)$$

Now let $\epsilon \to 0$. We have

$$C(S') - C(S) \leqslant \log\left[\frac{v_1(S')}{v_1(S)}(1 + \delta)^n\right]$$

Since $S' \subset (1 + \delta)S$, $v_1(S') \leqslant (1 + \delta)^n v_1(S)$. Hence

$$C(S') - C(S) \leqslant 2n \log(1 + \delta)$$

The same argument applies with S and S' interchanged. Therefore

$$\rho(S, S') < \alpha \Rightarrow |C(S') - C(S)| \leqslant 2n \log(1 + \delta)$$

This states the continuity of $C(S)$ in the Hausdorff metric.

We now consider the problem of when $C(S)$ can be 0. If translates of S tile E^n, then the unit cube can be covered by translates of $\frac{1}{2}\epsilon S$ with disjoint interiors, for arbitrarily small ϵ. Let $W = \{W_j\}$ be such a covering, and $W^* = \{W_j^*\}$ its restriction to the unit cube. Counting only sets of W which intersect the unit cube, these lie in a cube of side $1 + 4\epsilon K$, where K is the largest value of any coordinate in the unit sphere $\| x \|_S \leqslant 1$. Hence

$$\begin{aligned} O \leqslant D(\epsilon) &\leqslant \sum m(W_j^*) \log[1/m(W_j^*)] - \log(2^n/\epsilon^n v_1) \\ &\leqslant \sum m(W_j) \log[1/m(W_j)] - \log(2^n/\epsilon^n v_1) \\ &\leqslant [(1 + 4\epsilon K)^n - 1] \log(2^n/\epsilon^n v_1) \end{aligned}$$

for $\epsilon^n v_1 < 1/e$. Taking the limit as $\epsilon \to 0$, $C(S) = 0$.

Now suppose $C(S) = 0$. Then, for $L = 1$ and given ϵ we have for the partition V of (6)

$$\begin{aligned} \log(2^n/\epsilon^n v_1) &\leqslant \sum m(V_j) \log[1/m(V_j)] \\ &= \log(2^n/\epsilon^n v_1) + g_2(\epsilon) \end{aligned}$$

Given $\delta > 0$, let K be the set of indices j for which $m(V_j) < (1 + \delta)^{-1} 2^{-n} \epsilon^n v_1$. Then

$$\sum_K m(V_j) \log(1 + \delta) \leqslant g_2(\epsilon) \tag{16}$$

Let $\epsilon = 1/q$, q a positive integer, and partition the cube (4) into q^n cubes C_l, $1 \leqslant l \leqslant q^n$, of side $1/q$. Let $V_{j,l} = V_j \cap C_l$. Then from (16),

$$\sum_{l=1}^{q^n} \left(\sum_K m(V_{j,l}) \right) \leqslant g_2(1/q)/\log(1 + \delta)$$

Hence there is an index $l = r$ for which

$$\sum_K m(V_{j,r}) \leqslant q^{-n} g_2(1/q)/\log(1 + \delta) \tag{17}$$

Suppose that C_r is the cube

$$a_k < x_k < a_k + 1/q, \qquad k = 1,\dots, n$$

Denote its vertex $(a_1, \dots, a_n)$ by a. Let

$$\mathscr{W}^{(q)} = \{W_j^{(q)} = \{q(V_j - a)/V_j \in \mathscr{V}, j \notin K\}$$

The sets of $\mathscr{W}^{(q)}$ have diameter $\leqslant 1$ and measures $\geqslant (1 + \delta)^{-1} 2^{-n} v_1$. By (17), the part of the unit cube not covered by $\mathscr{W}^{(q)}$ has measure at most $g_2(1/q)/\log(1 + \delta)$.

Let $\mathscr{W}^{(q)*}$ consist of the closures of the sets of $\mathscr{W}^{(q)}$ which intersect the unit cube. From the lower bound on the measures of these sets and the bound on their diameters, the number of sets in $\mathscr{W}^{(q)*}$ is bounded, independent of q. Hence there is a number m and a sequence $\{q_s\}$ of values of q for which $\mathscr{W}^{(q)*}$ contains m sets:

$$\mathscr{W}^{(q_s)^*} = \{W_j^{(q_s)^*}, j = 1,\dots, m\}$$

where the sets are indexed in any order. Since the compact subsets of E^n form a locally compact space in the Hausdorff metric, there is a subsequence T of $\{q_s\}$ such that

$$W_j = \rho\text{-}\lim_{q \in T} W_j^{(q)*}$$

exists for $j = 1,\ldots, m$. It is easily shown that the W_j are sets with diameters $\leqslant 1$ and measures $\geqslant (1 + \delta)^{-1} 2^{-n} v_1$, the W_j have disjoint interiors, and the part of the cube not covered by $\{W_j\}$ has measure $0 = \lim g_2(1/q)/\log(1 + \delta)$.

Thus, for each $\delta > 0$, there is such a collection $\mathscr{W}_\delta = \{W_j\}$ of closed sets which covers the unit cube. By taking the limit on a sequence $\delta_j \to 0$, we get a covering of the unit cube by closed sets with diameters $\leqslant 1$, measures $\geqslant 2^{-n} v_1$, and disjoint interiors. These sets must be translates of $\frac{1}{2}S$. Hence the unit cube is tiled by translates of $\frac{1}{2}S$.

Similarly, starting with any $L > 0$, we get a tiling of the cube (4) by translates of $\frac{1}{2}S$. Translate (4) and its tiling to the position with the origin centered in the cube and multiply all coordinates by 2. Then, as $L \to \infty$, we have a sequence of expanding cubes whose union is E^n, each tiled by translates of S. Again taking the limit appropriately, we get a tiling of E^n by translates of S. This completes the proof that $C(S) = 0$ if and only if such a tiling exists.

To prove that $C(S) \leqslant 1$, we use Sec. 6 of Reference 2. Let βv_1 be the minimum measure of the 2^n pieces into which S is cut by the coordinates hyperplanes. Then for small ϵ, if $S_{\epsilon/2}(x)$ denotes the translate of $\frac{1}{2}\epsilon S$ centered at x, interested with X, we have

$$m(S_{\epsilon/2}(x) \geqslant \beta 2^{-n} \epsilon^n v_1, \qquad x \in X,$$

while

$$m(S_{\epsilon/2}(x)) = 2^{-n} \epsilon^n v_1$$

except on a subset of X near the boundary of measure $O(\epsilon)$. The above-cited reference yields

$$\begin{aligned} H_\epsilon(X) \leqslant & \int \log \frac{1}{m[S_{\epsilon/2}(x)]}\, dm \\ & + \Big[\int \log \frac{1}{1 - m[S_{\epsilon/2}(x)]}\, dm\Big]\Big[\int \frac{1 - m[S_{\epsilon/2}(x)]}{m[S_{\epsilon/2}(x)]}\, dm\Big] \end{aligned} \tag{18}$$

so that

$$\begin{aligned} H_\epsilon(X) &\leqslant [1 + O(\epsilon)] \log \frac{2^n}{\epsilon^n v_1} + \Big[[1 + O(\epsilon)] \frac{\epsilon^n v_1}{2^n}\Big]\Big[[1 + O(\epsilon)] \frac{2^n}{\epsilon^n v_1}\Big] \\ &= \log \frac{2^n}{\epsilon^n v_1} + 1 + O\left(\epsilon \log \frac{1}{\epsilon}\right) \end{aligned}$$

Then from the definition (7), we find

$$D(\epsilon) \leqslant 1 + O\left(\epsilon \log \frac{1}{\epsilon}\right) \tag{19}$$

From (11) we find

$$C(S) \leqslant 1 \tag{20}$$

as required. This completes the proof of Theorem 1.

Remark. The so-called "deterministic case" of (7) defines a $D'(\epsilon)$ as

$$D'(\epsilon) = H_\epsilon'(X) - \log(2^n/\epsilon^n v_1)$$

where $H_\epsilon'(X)$ is the epsilon entropy of the compact metric space X, that is, the minimum of the logarithm of the number of sets in an ϵ-covering of all of X. We can prove in the same way that

$$\lim_{\epsilon \to 0} D'(\epsilon) = C'(S)$$

the *deterministic packing constant* of S, exists. Also, $\infty > C'(S) \geqslant C(S)$, and $C'(S) = 0$ if and only if translates of S tile E^n. However, it is not true that $C(S)$ is uniformly bounded in n. In fact, Theorem 3.2 of Reference 6 shows in effect that

$$C'(S) \leqslant [1 + o(1)] \log n$$

and, for S the n-ball (Theorem 8.1 of Reference 6),

$$C'(S) \geqslant [1 - o(1)] \log n$$

Thus, $C'(S)$ can be arbitrarily large, even though $C(S) \leqslant 1$. What this means is that compact convex symmetric sets in E^n with nonempty interior pack vastly better if one is allowed to weight sets according to their measure instead of counting how many are necessary. For the n-ball of radius $\epsilon/2$, a lot of pieces of very small measure must be used to partition the unit n-cube; if one did not have to cover everything, but only most of the cube, a lot fewer sets would be needed. The difference

$$C'(S) - C(S)$$

is a measure of the extra packing difficulty one has in packing S when the sizes of the additional pieces cannot be taken into account.

3. INTRODUCTION OF DIFFERENTIAL ENTROPY

We call a function $f(x)$ defined on E_n *strongly integrable* if $f \in L_1(E_n)$, and its integral is approximated by Riemann-type sums over partitions of E_n of small mesh: $f(x)$ is strongly integrable if for any $\eta > 0$ there is a $\delta > 0$ such that if $U = \{U_j\}$ is any δ-partition of E_n,

$$\left| \int_{E_n} f(x)\, dm(x) - \sum m(U_j) f(\xi_j) \right| < \eta \tag{21}$$

where $\{\xi_j\}$ is any sequence of points with $\xi_j \in U_j$. This condition is satisfied, for example, if $f(x)$ is Riemann-integrable over any bounded region in E_n, and $f(x) \to 0$ rapidly at infinity.

We say that $f(x)$ is *strongly integrable of order* α if there is a constant A such that (21) is true for all sufficiently small δ, if $\eta = A\delta^\alpha$. This condition is satisfied, for example, if $f(x)$ satisfies an inequality of the form

$$|f(x) - f(x')| < B|x - x'|^{\alpha/n} g(x), \qquad B > 0$$

where $g(x)$ is a strongly integrable function.

Define the differential entropy $H(p)$ of a probability density function p on E^n as the possibly negative infinite integral

$$H(p) = \int p \log(1/p)\, dm$$

Using this concept, the following theorem can be stated and proved.

Theorem 2. Let $X = \{(x_1, ..., x_n)\}$ be a real normed linear space of dimension n arising from a compact convex symmetric set S with nonempty interior, together with a Borel probability distribution μ with a density $p(x)$. If $p(x)$ is continuous and there is an $\alpha > 0$ such that $p(x)$ and $p(x)\log[1/p(x)]$ are strongly integrable of order α, then

$$H_\epsilon(X) = n \log(2/\epsilon) + H(p) + \log(1/v_1) + C(S) + o(1) \tag{22}$$

as $\epsilon \to 0$, where v_1 is the Lebesgue measure of S, and $C(S)$ is the entropic packing constant of S.

Proof. Let $U = \{U_j\}$ be any ϵ-partition of X. We have

$$\begin{aligned} H(U) &= \sum \mu(U_j) \log[1/\mu(U_j)] \\ &= \sum m(U_j)\, p(\xi_j) \log[1/m(U_j)p(\xi_j)] \\ &= \sum m(U_j)\, p(\xi_j) \log[1/p(\xi_j)] + \sum p(\xi_j)\, m(U_j) \log[1/m(U_j)] \\ &= H_1(U) + H_2(U), \quad \text{say,} \end{aligned}$$

where ξ_j is the point of U_j at which $p(x)$ takes its average value in U_j. By hypothesis, there is a constant A_1 such that

$$|H_1(U) - H(p)| < A_1\, \epsilon^\alpha \tag{23}$$

Take $\delta = \sqrt{\epsilon}$, and partition X into coordinate cubes of side δ by the hyperplanes $x_k = j\delta$, $-\infty < j < \infty$, $k = 1, ..., n$. For ϵ sufficiently small, all the terms in the series for $H_2(U)$ are nonnegative. Let the cubes of side δ be $\{K_r\}$:

$$H_2(U) \geqslant \sum_r \sum_{U_j \subset K_r} p(\xi_j)\, m(U_j) \log \frac{1}{m(U_j)} \tag{24}$$

Let $\bar{p}_r$ be the minimum value of $p(x)$ in K_r. Then by Eq. (5) of Theorem 1,

$$H_2(U) \geqslant \sum_r \bar{p}_r\, \delta^n[\log(2^n/\epsilon^n v_1) + C(S) + g_1(\sqrt{\epsilon})] \tag{25}$$

By hypothesis, there is a constant A_2 such that

$$\left| \sum_r \bar{p}_r \, \delta^n - 1 \right| < A_2 \epsilon^{\alpha/2}$$

The expression in brackets in (25) is $o(\epsilon^{-\alpha/2})$ as $\epsilon \to 0$. Hence

$$H_2(U) \geqslant \log(2^n/\epsilon^n v_1) + C(S) + g_1^*(\epsilon) \tag{26}$$

where $g_1^*(\epsilon) \to 0$ as $\epsilon \to 0$.

For a special choice of U, take the partition of Theorem 1 (with $L = \delta$), together with its translation into all the other cubes of $\{K_r\}$. Then equality holds in (24), and instead of (25) we have

$$H_2(U) \leqslant \sum_r p_r^* \, \delta^n [\log(2^n/\epsilon^n v_1) + C(S) + g_2(\sqrt{\epsilon})]$$

if p_r^* is the maximum of $p(x)$ in K_r. This leads as above to

$$H_2(U) \leqslant \log(2^n/\epsilon^n v_1) + C(S) + g_2^*(\epsilon) \tag{27}$$

where $g_2^*(0^+) = 0$.

By (23) and (26),

$$H_\epsilon(X) \geqslant n \log(2/\epsilon) + H(p) + \log(1/v_1) + C(S) + g_1^*(\epsilon) - A_1 \epsilon^\alpha$$

Using the special partition for which (27) holds, we get

$$H_\epsilon(X) \leqslant n \log(2/\epsilon) + H(p) + \log(1/v_1) + C(S) + g_2(\epsilon) + A_1 \epsilon^\alpha$$

Hence (22) is true. Theorem 2 is proved.

What Theorem 2 means is that the differential entropy $H(p)$ for a nice density p is, except for a term approaching 0 with ϵ, the difference between the ϵ-entropy (of the space with metric obtained from $\| \cdots \|_S$ and probability from p) and the logarithm of the reciprocal of the volume of the sphere of diameter ϵ in the norm, less a term $C(S)$ that measures how badly S fails to close-pack all of n-space. For S the unit cube, $H(p)$ is just the difference of the epsilon entropy of the space and the logarithm of the reciprocal of the volume of a sphere of diameter ϵ in that norm (the so-called sup norm or L_∞ norm). This is one explanation of the term "differential entropy."

Counterexample. The condition that p be continuous and strongly integrable over E_n cannot be relaxed. To show this, let $\{p_i\}$ be the sequence given by

$$p_i = \frac{c}{i \log^2(i+1)}, \qquad i \geqslant 1$$

where

$$c^{-1} = \sum_{i=1}^{\infty} \frac{1}{i \log^2(i+1)}$$

Let p be the indicator function of the set A, where A is the union of the intervals $[i, i + p_i]$, $i \geqslant 1$. Then $p \log(1/p)$ is identically zero, *a fortiori* strongly integrable of order 1. And yet $H_\epsilon(X)$ is infinite for $\epsilon > 0$. This example can be modified so that p is continuous but not strongly integrable, keeping the strong integrability of $p \log(1/p)$.

REFERENCES

1. E. C. Posner, E. R. Rodemich, and H. Rumsey, Jr., "Epsilon entropy of stochastic processes," *Ann. Math. Stat.* **38**:1000–1020 (1967).
2. E. C. Posner and E. R. Rodemich, "Epsilon entropy and data compression," in preparation.
3. C. E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* **27**:379–423 (1948).
4. F. Hausdorff, *Set Theory* (trans. from German), Chelsea, New York (1957).
5. L. Fejes Tóth, *Regular Figures*, Pergamon Press, New York (1964).
6. C. A. Rogers, *Packing and Covering*, Cambridge University Press (1964).