ENTROPIES OF SETS OF FUNCTIONS OF BOUNDED VARIATION

G. F. CLEMENTS

1. Introduction. In this paper the entropies of several sets of functions of bounded variation are calculated. The entropy of a metric set, a notion first introduced by Kolmogorov in **(2)**, is a measure of its size in terms of the minimal number of sets of diameter not exceeding 2ϵ necessary to cover it. Using this notion, Kolmogorov **(4**; p. 357**)** and Vituškin **(7)** have shown that not all functions of n variables can be represented by functions of fewer variables if only functions satisfying certain smoothness conditions are allowed. For an exposition of this application and other results concerning entropy the reader is referred to the paper of G. G. Lorentz **(5)**. Before stating our results, we first collect the basic facts and definitions **(4**, p. 279**)**. Let A be a non-void subset of a metric space W.

DEFINITION 1. *A system γ of sets $U \subseteq W$ is called an ϵ-cover of A if for each U in γ, the diameter of U, $d(U)$, does not exceed 2ϵ, and*

$$A \subseteq \bigcup_{U \epsilon \gamma} U.$$

DEFINITION 2. *A set $U \subseteq W$ is an ϵ-net for A if each point of A has distance not exceeding ϵ from at least one point of U.*

DEFINITION 3. *A set $U \subseteq W$ is said to be ϵ-distinguishable if the distance between any two points of U is greater than ϵ.*

In what follows we shall deal exclusively with *totally bounded* sets; that is, sets having a finite ϵ-cover for each $\epsilon > 0$, or, equivalently, sets having a finite ϵ-net for each $\epsilon > 0$, or sets for which each ϵ-distinguishable subset is finite. In particular, compact sets are totally bounded. We are interested in the following functions:

$N_\epsilon^W(A)$, the minimal number of points in W which form an ϵ-net for A;

$N_\epsilon(A)$, the minimal number of sets in an ϵ-cover of A;

$M_\epsilon(A)$, the maximal number of points in an ϵ-distinguishable subset of A.

The dyadic logarithms of $N_\epsilon(A)$ and $M_\epsilon(A)$ are called the *entropy* and the *capacity* of A and are denoted by $H_\epsilon(A)$ and $C_\epsilon(A)$ respectively:

$$H_\epsilon(A) = \log N_\epsilon(A), \qquad C_\epsilon(A) = \log M_\epsilon(A).$$

Received May 24, 1962. This research is part of the author's Ph.D. thesis. It has been supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF4 9(638)-619.

It is unusual to be able to determine these functions exactly and one is usually content with finding their order. We write $f(\epsilon) \prec g(\epsilon)$ for $f(\epsilon) = O(g(\epsilon))$ and $f(\epsilon) \approx g(\epsilon)$ if both $f(\epsilon) = O(g(\epsilon))$ and $g(\epsilon) = O(f(\epsilon))$. Thus for various sets A we seek a function $h(\epsilon)$ for which $H_\epsilon(A) \approx h(\epsilon)$ holds. The basic tool to this end is the following theorem (**4**, p. 282).

THEOREM. *For each totally bounded set A of a metric space W, the inequalities*

$$M_{2\epsilon}(A) \leqslant N_\epsilon(A) \leqslant N_\epsilon^W(A) \leqslant M_\epsilon(A) \tag{1}$$

hold, and therefore

$$C_{2\epsilon}(A) \leqslant H_\epsilon(A) \leqslant C_\epsilon(A).$$

In Section 2 we consider continuous functions $f(x)$ defined on $[0, 1]$ with $|f(x)| \leqslant M$ and total variation over $[0, 1]$ not exceeding some positive constant B not depending on f. Under the uniform metric ρ, defined by

$$\rho(f, g) = \max_{x\in[0,1]} |f(x) - g(x)|,$$

this set is not totally bounded: the functions $y = nx$ truncated at $y = 1$, $n = 1, 2, \ldots$ contain a non-finite $\frac{1}{2}$-distinguishable set. To get a totally bounded set so that the entropy will exist, we further require that f satisfy a Lipschitz condition of order α, $0 < \alpha < 1$:

$$|f(x') - f(x'')| \leqslant |x' - x''|^\alpha \qquad \text{for } x', x'' \in [0, 1]]. \tag{2}$$

Calling this set V_α, we shall show that $H_\epsilon(V_\alpha) \approx (1/\epsilon)\log(1/\epsilon)$. For the sake of comparison, the set of all functions defined on $[0, 1]$ with $|f(x)| \leqslant M$ which satisfy (2), with $0 < \alpha \leqslant 1$, has entropy of order $(1/\epsilon)^{1/\alpha}$ (**4**, p. 308).

In Section 3, we again consider sets of functions of bounded variation, but take for the distance between two functions the Hausdorff distance between their graphs (**1**, p. 166). This gives a smaller metric than ρ and the sets we consider are totally bounded even without the assumption (2). For the set V_L which consists of functions f defined on $[0, 1]$ which satisfy $|f(x)| \leqslant M$ and

$$\operatorname*{Var}_{[0,1]} f \leqslant L,$$

and for the set ${}_LX_n^*$ which consists of continuous curves of length not exceeding L contained in a n-dimensional cube, we find that $H_\epsilon \approx (1/\epsilon)$. We also show, for the set C_B of functions f defined on $[0, 1]$ for which $f(x+)$ and $f(x-)$ exist and $|f(x)| \leqslant B$ for all $x \in [0, 1]$, that $H_\epsilon(C_B) \approx (1/\epsilon)\log(1/\epsilon)$.

2. The entropy of V_α.

With V_α as defined in the introduction, we now prove the following theorem.

THEOREM 1. $H_\epsilon(V_\alpha) \approx C_\epsilon(V_\alpha) \approx (1/\epsilon)\log(1/\epsilon)$.

Proof. By constructing a 3ϵ-net for V_α we shall first show that

$$H_{3\epsilon}(V_\alpha) \prec (1/\epsilon)\log(1/\epsilon),$$

or equivalently, $H_\epsilon(V_\alpha) \prec (1/\epsilon)\log(1/\epsilon)$. To this end, let $\epsilon > 0$ be given. Take

$$n = n_\epsilon = [\epsilon^{-1/\alpha}] + 1 \approx \epsilon^{-1/\alpha}, \qquad \text{and} \qquad \delta = (1/n) \approx \epsilon^{1/\alpha}.$$

Here and below $[x]$ denotes the largest integer not exceeding x. Put $x_i = i\delta$, $i = 0, 1, \ldots, n$, and $l = \log(1/\epsilon)$. Denote by G the smallest integer for which $Gl \geqslant n$, and let $q_i = x_{(i-1)l}$, $i = 1, \ldots, G$. Let $I_i = [q_i, q_{i+1})$, $i = 1, \ldots, G-1$, and $I_G = [q_G, 1]$. In this way $[0, 1]$ is divided into $G < (n/l) + 1 < 2n/l$ intervals I_i all of which, except I_G, consist of l consecutive subintervals of the form $[x_{r-1}, x_r)$. Corresponding to a given $f \in V_\alpha$, we shall call I_i a good or bad interval according as the variation of f on I_i does not exceed $\delta^\alpha = \epsilon_0$ or does exceed ϵ_0.

For $f \in V_\alpha$, we define a function $c(x) = c_f(x)$ on $[0, 1]$ in such a way that $\rho(f, c_f) \leqslant 3\epsilon_0 < 3\epsilon$; $\{c_f\}_{f\epsilon V_\alpha}$ will therefore be a 3ϵ-net for V_α. As we define c_f, we also explain how to label it with a matrix M_f having G rows. The first i rows of this matrix will determine $c_f(x)$ over $I_1, I_2, \ldots, I_i$ and conversely. These labels, in a 1–1 correspondence with the functions c_f, $f \in V_\alpha$, will be helpful in estimating the number of functions in our 3ϵ-net.

The matrix M_f and the function $c(x) = c_f(x)$ are defined in the following way. On the interval I_1, define $c(x)$ by

$$c(x) = [f(x_{i-1})/\epsilon_0]\epsilon_0, \qquad x \in [x_{i-1}, x_i), \qquad i = 1, 2, \ldots, l$$

and take the first row of M_f to be $[f(x_0)/\epsilon_0], [f(x_1)/\epsilon_0], \ldots, [f(x_{l-1})/\epsilon_0]$. If $c(x)$ is defined on $I_1, I_2, \ldots, I_{k-1}$ and the first $(k-1)$ rows of M_f are defined, define $c(x)$ on $I_k = [q_k, q_{k+1})$ and the kth row of M_f according to the following rules:

1. If I_k is a bad interval, define

$$c(x) = [f(x_{i-1})/\epsilon_0]\epsilon_0 \quad \text{for } x \in [x_{i-1}, x_i), \quad i = (k-1)l + 1, \ldots, kl,$$

and take the kth row of M_f to be

$$(3) \qquad [f(x_{(k-1)l})/\epsilon_0], \qquad [f(x_{(k-1)l+1})/\epsilon_0], \ldots, [f(x_{kl-1})/\epsilon_0].$$

2. If I_{k-1} is a bad interval and I_k is a good interval, define

$$c(x) = [f(q_k)/\epsilon_0]\epsilon_0 \qquad \text{for } x \in I_k,$$

and take the kth row of M_f to be b ; $[f(q_k)/\epsilon_0]$.

3. If I_{k-1} and I_k are good intervals, and there exists a point $x^* \in I_k$ such that $|c(q_{k-1}) - f(x^*)| > 3\epsilon_0$, define

$$c(x) = [f(q_k)/\epsilon_0]\epsilon_0 \qquad \text{for } x \in I_k,$$

and take the kth row of M_f to be g ; $[f(q_k)/\epsilon_0]$.

https://doi.org/10.4153/CJM-1963-045-3 Published online by Cambridge University Press

4. If I_{k-1} and I_k are good intervals, and $|c(q_{k-1}) - f(x)| \leqslant 3\epsilon_0$ for $x \in I_k$, define

$$c(x) = c(q_{k-1}) \qquad \text{for } x \in I_k,$$

and take the kth row of M_f to be 0.

Thus corresponding to $f \in V_\alpha$, we have defined $c(x) = c_f(x)$ on $[0, 1]$. We now show that $|f(x) - c_f(x)| \leqslant 3\epsilon_0$ for $x \in [0, 1]$, from which it follows that $\rho(f, c_f) \leqslant 3\epsilon_0 < 3\epsilon$. Let $x \in [0, 1]$ be given, and take k and r such that $x \in [x_{r-1}, x_r) \subset I_k$. We shall speak of I_k as being of type i if rule i was used to define $c(x)$ over I_k, $i = 1, 2, 3, 4$, and shall consider I_1 to be of type 1.

If I_k is of type 1, then $c(x) = [f(x_{r-1})/\epsilon_0]\epsilon_0$ for $x \in [x_{r-1}, x_r)$, and so

$$(4) \qquad \begin{aligned} |f(x) - c(x)| &\leqslant |f(x) - f(x_{r-1})| + |f(x_{r-1}) - c(x)| \\ &\leqslant |x - x_{r-1}|^\alpha + \epsilon_0 < \delta^\alpha + \epsilon_0 = 2\epsilon_0. \end{aligned}$$

If I_k is of type 2 or 3, then, since the variation of f on I_k does not exceed ϵ_0, we have for $x \in I_k$

$$(5) \qquad \begin{aligned} |f(x) - c(x)| &= |f(x) - [f(q_k)/\epsilon_0]\epsilon_0| \\ &\leqslant |f(x) - f(q_k)| + \epsilon_0|f(q_k)/\epsilon_0 - [f(q_k)/\epsilon_0]| < \epsilon_0 + \epsilon_0 = 2\epsilon_0. \end{aligned}$$

It follows from (5) that $|f(x) - c(q_k)| \leqslant 2\epsilon_0$. In particular, for each interval I_k of type 3,

$$(6) \qquad c(q_{k-1}) \neq c(q_k).$$

Finally, if I_k is of type 4, the very criterion for applying rule 4 shows that $|c(x) - f(x)| \leqslant 3\epsilon_0$ for $x \in I_k$.

Thus $\{c_f(x)\}_{f \in V_\alpha}$ is a 3ϵ-net for V_α. To count how many distinct functions there are in this net, it suffices to count how many distinct matrices M_f there are, since the matrices M_f and the functions $c_f(x)$ for $f \in V_\alpha$ correspond in a 1–1 way. We shall refer to the kth row of M_f as being of type i if I_k is of type i, $i = 1, 2, 3, 4$. Let P_1, P_2, P_3 denote the numbers of different ways in which rows of type 1, 2, 3, correspondingly, can occur in M_f. The knowledge of all rows of these three types of M_f determines the matrix M_f completely; hence there are at most $P_1P_2P_3$ different matrices M_f.

We shall estimate the numbers P_i from above. We begin by remarking that the Lipschitz condition (2) gives

$$\begin{aligned} |f(x_i)/\epsilon_0 - [f(x_{i-1})/\epsilon_0]| &\leqslant 1 + |f(x_i)/\epsilon_0 - f(x_{i-1})/\epsilon_0| \\ &\leqslant 1 + (1/\epsilon_0)|x_i - x_{i-1}|^\alpha \leqslant 2. \end{aligned}$$

It follows that $[f(x_i)/\epsilon_0]$ is one of the three integers

$$[f(x_{i-1})/\epsilon_0] - 1, \qquad [f(x_{i-1})/\epsilon_0], \qquad [f(x_{i-1})/\epsilon_0] + 1.$$

The first term of a type 1 row (3) can take at most $2[B/\epsilon_0] + 1$ values, since all values of $|f(x)|$ do not exceed B. For each following term, there are at most three possibilities. Hence there exist at most

https://doi.org/10.4153/CJM-1963-045-3 Published online by Cambridge University Press

$$(2[B/\epsilon_0] + 1)3^l$$

type 1 rows.

Since the variation of f on a type 1 interval I_k other than I_1 exceeds ϵ_0, and its total variation is at most B, there are at most $[B/\epsilon_0] + 1$ type 1 rows. In a matrix of G rows, there are

$$J = \sum_{i=0}^{[B/\epsilon_0]+1} \binom{G}{i} \leqslant ([B/\epsilon_0] + 1)G^{[B/\epsilon_0]+1}$$

different selections of a subset of at most $[B/\epsilon_0] + 1$ rows. Hence the number of ways in which rows of type 1 can occur in M_f does not exceed

$$P_1 \leqslant J\{(2B/\epsilon_0 + 1)3^l\}^{B/\epsilon_0+1}. \tag{7}$$

Similarly, since there is a bad interval preceding each type 2 interval, there are not more than B/ϵ_0 type 2 rows, and not more than J orders of type 2 rows. Then, since $[f(q_k)/\epsilon_0]$ takes at most $2[B/\epsilon_0] + 1$ values, we get

$$P_2 \leqslant J(2B/\epsilon_0 + 1)^{B/\epsilon_0}. \tag{8}$$

Finally, we show that the number of type 3 rows in M_f does not exceed $[B/\epsilon_0]$. To prove this, we associate with each type 3 interval I_k a chain

$$I_{k-m}, I_{k-m+1}, \ldots, I_k \qquad (m \geqslant 1) \tag{9}$$

such that the variation of f on the union

$$U = I_{k-m+1} \cup I_{k-m+2} \cup \ldots \cup I_k$$

exceeds ϵ_0 and such that I_k is the only type 3 interval in U. We take m to be the largest integer with the property that all intervals in the chain (9) are good and that

$$c(q_{k-m}) = c(q_{k-m+1}) = \ldots = c(q_{k-1}). \tag{10}$$

The chain is unextendable either because I_{k-m-1} is bad or because I_{k-m} is of type 3. In both cases (see (4), (5)) we have for each point $x \in I_{k-m}$,

$$|f(x) - c(q_{k-m})| \leqslant 2\epsilon_0.$$

On the other hand, for some $x^* \in I_k$, $|f(x^*) - c(q_{k-1})| > 3\epsilon_0$. From this and (10) it follows that $|f(x) - f(x^*)| > \epsilon_0$, for $x \in I_{k-m}$, and making $x \to q_{k-m+1}$,

$$|f(q_{k-m+1}) - f(x^*)| \geqslant \epsilon_0,$$

so that the variation of f on U is at least ϵ_0. It is also clear, in view of (6) and (10), that I_k is the only type 3 interval in U. Hence we again obtain

$$P_3 \leqslant J(2B/\epsilon_0 + 1)^{B/\epsilon_0}. \tag{11}$$

Taking logarithms in the relation $N_{3\epsilon}(V_\alpha) \leqslant P_1P_2P_3$ and using (7), (8), and (11) we obtain

$$H_{3\epsilon}(V_\alpha) = \log N_{3\epsilon}(V_\alpha) \prec \log J + (B/\epsilon_0) \log (2B/\epsilon_0) \\ \prec (1/\epsilon_0) \log (1/\epsilon_0) + (1/\epsilon_0) \log G,$$

and since $G \leqslant 2n/l = 2\epsilon^{-1/\alpha} [\log (1/\epsilon)]$, we obtain

$$H_{3\epsilon}(V_\alpha) \prec (1/\epsilon) \log (1/\epsilon). \tag{12}$$

To estimate $H_\epsilon(V_\alpha)$ from below, partition $[0, 1]$ by points $x_i = i\delta$, $i = 0, 1, \ldots, n$, where $n = [\epsilon^{-1/\alpha}] - 1 < \epsilon^{-1/\alpha}$ and $\delta = 1/n > \epsilon^{1/\alpha}$. With $G = [B/2\delta^\alpha]$ and $l = [n/G] \approx \epsilon^{1-1/\alpha}$, let $I_i = [x_{(i-1)l}, x_{il})$, $i = 1, 2, \ldots, G - 1$, and $I_G = [x_{(G-1)l}, 1]$. We have thus separated $[0, 1]$ into G intervals I_i, $i = 1, 2, \ldots, G$, each interval consisting of l subintervals of the form $[x_{r-1}, x_r]$ except for I_G which may contain more than l subintervals. Now consider all functions f which are 0 on $[0, 1]$ except for two consecutive subintervals in each interval I_i over which the graph of f along with the x axis forms an isosceles triangle of altitude $\delta^\alpha > \epsilon$. The variation of f over $[0, 1]$ is exactly $2G\delta^\alpha \leqslant B$, f satisfies the Lipschitz condition of order α, and f is therefore in V_α. Distinct functions of this type differ by $\delta^\alpha > \epsilon$ at some x_i, so this set of functions is ϵ-distinguishable. If the number of these functions is $K(\epsilon)$, we have

$$M_\epsilon(V_\alpha) \geqslant K(\epsilon) \succ (l/2)^G \succ (\epsilon^{1/\alpha - 1})^{[B/2\delta^\alpha]},$$

and therefore

$$H_{\epsilon/2}(V_\alpha) \succ \log M_\epsilon(V_\alpha) \succ (1/\epsilon) \log (1/\epsilon), \quad \text{or} \quad H_\epsilon(V_\alpha) \succ (1/\epsilon) \log (1/\epsilon),$$

which proves the theorem.

3. Entropies of sets of functions of bounded variation in the Hausdorff metric. Let (X, η) be a totally bounded metric space with metric η. Let X^* be the set of all non-void closed subsets of (X, η). For A, B in X^*, we define (**1**, p. 166)

$$\sigma(A, B) = \inf \{\epsilon \mid S_\epsilon(A) \supset B \text{ and } S_\epsilon(B) \supset A\},$$

where

$$S_\epsilon(A) = \bigcup_{a \in A} \{y \mid \eta(y, a) < \epsilon\}.$$

Then σ is a metric on X^* (the Hausdorff metric).

It is straightforward to check that if $F : x_1, x_2, \ldots, x_n$ is a maximal ϵ-distinguishable set in (X, η), then the $2^n - 1$ non-void subsets of F are ϵ-distinguishable in (X^*, σ) and form a 2ϵ-net in (X^*, σ). From this and (1), it follows that

$$\tfrac{1}{2} 2^{N_{2\epsilon}(X,\eta)} \leqslant 2^{N_{2\epsilon}(X,\eta)} - 1 \leqslant 2^{M_{2\epsilon}(X,\eta)} - 1 \leqslant M_{2\epsilon}(X^*, \sigma) \\ \leqslant N_\epsilon(X^*, \sigma) \leqslant N_\epsilon^{(X^*,\sigma)}(X^*, \sigma) \leqslant 2^{N_{\epsilon/2}(X,\eta)} - 1 \leqslant 2^{N_{\epsilon/2}(X,\eta)},$$

and, taking logarithms,

$$N_{2\epsilon}(X, \eta) - 1 \leqslant C_{2\epsilon}(X^*, \sigma) \leqslant H_\epsilon(X^*, \sigma) \leqslant N_{\epsilon/2}(X, \eta). \tag{13}$$

If X_n is the cube $\{(x_1, x_2, \ldots, x_n) \mid 0 \leqslant x_i \leqslant M,\ i = 1, 2, \ldots, n\}$ in the n-dimensional Euclidean space with the usual metric, then $N_\epsilon(X_n) \approx (1/\epsilon)^n$ (**4**, p. 300). Hence from (13) follows:

THEOREM 2. $H_\epsilon(X_n^*, \sigma) \approx C_\epsilon(X_n^*, \sigma) \approx (1/\epsilon)^n$.

We now consider subsets of X_n^*. By the "curve C given parametrically by the co-ordinate functions $x_1(t), x_2(t), \ldots, x_n(t)$, $a \leqslant t \leqslant b$," we mean the set

$$C = \{(x_1, x_2, \ldots, x_n) \mid x_i = x_i(t), t \in [a, b], i = 1, 2, \ldots, n\}.$$

Letting ${}_cX_n^*$ $(n \geqslant 2)$ be the set of all such curves which are contained in X_n and have continuous co-ordinate functions, we have:

THEOREM 3. $H_\epsilon({}_cX_n^*) \approx C_\epsilon({}_cX_n^*) \approx (1/\epsilon)^n$.

Proof. Since ${}_cX_n^*$ is a subset of X_n^*, the estimate from above follows from the preceding theorem. To get the estimate from below, we exhibit a set of $2^{[M/\epsilon]^n} - 1$ curves in ${}_cX_n^*$ which are $\epsilon/2$-distinguishable. From this it will follow that

$$2^{[M/\epsilon]^n - 1} \leqslant 2^{[M/\epsilon]^n} - 1 \leqslant M_{\epsilon/2}({}_cX_n^*) \leqslant N_{\epsilon/4}({}_cX_n^*), \tag{14}$$

and hence

$$(1/\epsilon)^n \prec C_{2\epsilon}({}_cX_n^*) \leqslant H_\epsilon({}_cX_n^*),$$

proving the theorem.

To get the set of $\epsilon/2$-distinguishable curves, take $l = M/([M/\epsilon] - 1) > \epsilon$, and consider the set D consisting of the $[M/\epsilon]^n$ points x having co-ordinates $(k_1 l, k_2 l, \ldots, k_n l)$, where k_i is an integer, $0 \leqslant k_i \leqslant [M/\epsilon] - 1$, $i = 1, 2, \ldots, n$. Distinct subsets of D are at least l apart (in the Hausdorff metric). So by associating with each subset $x_1, x_2, \ldots, x_s$ of D a continuous curve passing through $x_1, x_2, \ldots, x_s$ and not approaching other points of D nearer than $l/2$, we get $2^{[M/\epsilon]^n} - 1$ curves mutually at least $l/2$ apart. Since $l > \epsilon$, these curves are $\epsilon/2$-distinguishable, which proves the theorem.

Now let ${}_LX_n^*$ $(n \geqslant 2)$ be the set of all curves C contained in X_n which can be represented parametrically in the form

$$C = \{(x_1, x_2, \ldots, x_n) \mid x_i = x_i(s), s \in [0, L], i = 1, 2, \ldots, n\}, \tag{*}$$

where the co-ordinate functions $x_i(s)$ are continuous and the parameter s is arc length. Any curve of length not exceeding L can be parametrized in this way. For instance, if

$$C = \{(x_1, x_2, \ldots, x_n) \mid x_i = x_i(s), s \in [0, L/2], i = 1, 2, \ldots, n\},$$

we also have

$$C = \{(x_1, x_2, \ldots, x_n) \mid x_i = \bar{x}_i(s), s \in [0, L], i = 1, 2, \ldots, n\},$$

https://doi.org/10.4153/CJM-1963-045-3 Published online by Cambridge University Press

where $\bar{x}_i(s) = x_i(s)$ for $s \in [0, L/2]$ and $\bar{x}_i(s) = x_i(L - s)$ for $s \in (L/2, L]$. The entropy of ${}_LX_n{}^*$ can be estimated from above as follows:

LEMMA 1. $H_\epsilon({}_LX_n{}^*) \prec 1/\epsilon$.

Proof. Each co-ordinate function $x_i(s)$ of a curve C of ${}_LX_n{}^*$ satisfies a Lipschitz condition of order 1:

$$|x_i(s') - x_i(s'')| \leqslant |s' - s''|.$$

Since the curves in ${}_LX_n{}^*$ are contained in X_n, we also have $0 \leqslant x_i(s) \leqslant M$. It is known (**4**, p. 308), that the set A of all functions $x(s)$ on $[0, L]$ which satisfy these two conditions has under the uniform metric ρ the entropy $H_\epsilon(A) \asymp 1/\epsilon$. If $A_i = A$, $i = 1, 2, \ldots, n$, then the product

$$P = \prod_{i=1}^{n} A_i$$

is a metric space with distance

$$\phi(x, x') = \sum_{i=1}^{n} \rho(x_i, x_i')$$

between the points $x = (x_1, \ldots, x_n)$ and $x' = (x_1', \ldots, x_n')$. Then (see **8**, p. 27, Lemma 1) $H_\epsilon(P) \leqslant nH_{\epsilon/n}(A)$. Now if a representation (*) is selected for each curve C in ${}_LX_n{}^*$, then ${}_LX_n{}^*$ is mapped into P, and if x, x' correspond to C, C', it is clear that $\sigma(C, C') \leqslant \phi(x, x')$. Then

$$H_\epsilon({}_LX_n{}^*) \leqslant H_\epsilon(P) \prec (1/\epsilon),$$

which proves the lemma.

To estimate $H_\epsilon({}_LX_n{}^*)$ from below, we consider a subset of ${}_LX_n{}^*$. Let M_L consist of curves C which can be given parametrically in the form (*), where $x_i(s) \equiv 0$, $3 \leqslant i \leqslant n$, and $x_1(s), x_2(s)$ are continuous monotonically increasing functions such that $x_1(0) = x_2(0) = 0$, $x_1(L) = 1$, and $x_2(L) = L - 1$. M_L is a subset of ${}_LX_n{}^*$ if we assume $M > L - 1$. The capacity of M_L may be estimated from below as follows:

LEMMA 2. $C_\epsilon(M_L) \succ 1/\epsilon$.

Proof. It suffices to consider the case $L = 2$. We construct a set of ϵ-distinguishable functions in M_2 in the following way. Put

$$n = n_\epsilon = [1/\epsilon] - 1 < 1/\epsilon, \ \delta = 1/n > \epsilon, \text{ and } x_i = i\delta, \ i = 0, 1, \ldots, n.$$

Consider all "step curves" $S = S_{m_1, m_2, \ldots, m_n}$ which consist of points (x, y) which satisfy for some k, $k = 1, \ldots, n$, the condition

$$x \in (x_{k-1}, x_k), y = m_k\delta; \qquad \text{or } x = x_k, m_k\delta \leqslant y \leqslant m_{k+1}\delta,$$

where m_0 is 0, m_{n+1} is n, and $m_1, m_2, \ldots, m_n$ is an increasing sequence of non-negative integers with $m_n = n$. These curves are in M_2. For convenience in counting, we label S by the sequence

$$s_k = m_k - m_{k-1}, \qquad k = 1, 2, \ldots, n, \tag{15}$$

of non-negative integers which add to n. The curves S correspond to their labels in a 1–1 way. Since $\delta > \epsilon$, it is easy to see that these step curves are ϵ-distinguishable (in the Hausdorff metric). To count them, consider the terms s_k in (15) which are strictly positive. If these terms are s_{k_j}, $j = 1, 2, \ldots, i$, we have

$$s_{k_1} + s_{k_2} + \ldots + s_{k_i} = n.$$

Such an ordered collection of integers is called a *composition of n of i parts.* For each composition, the $n - i$ zeros in $s_1, s_2, \ldots, s_i$ can be arranged in $\binom{n}{n-i}$ ways, and there are $\binom{n-1}{i-1}$ compositions of n having i parts (**6**, p. 124), so in all there are

$$F(n) = \sum_{i=1}^{n} \binom{n-1}{i-1}\binom{n}{n-i} = n^{-1}\sum_{i=1}^{n} i\binom{n}{i}^2 > n^{-1}[n/2]\binom{n}{[n/2]}^2$$

step curves. Using Stirling's formula, it is seen that

$$\log\binom{n}{[n/2]} \succ n,$$

so $\log F(n) \succ n$. Since $n = n_\epsilon \sim 1/\epsilon$, we have $\log F(n) \succ 1/\epsilon$, and

$$C_\epsilon(M_2) \geqslant \log F(n) \succ 1/\epsilon,$$

which proves the lemma.

Since M_L is a subset of ${}_LX_n{}^*$ (if $M > L + 1$), Lemmas 1 and 2 give:

THEOREM 4. *If R is ${}_LX_n{}^*$, M_L, or any subset of ${}_LX_n{}^*$ which contains M_L, then*

$$H_\epsilon(R) \approx C_\epsilon(R) \approx (1/\epsilon).$$

We now explain how the result $H_\epsilon(M_L) \approx 1/\epsilon$ can be interpreted as an entropy statement about a function space. If Q is any set of functions defined and bounded on $[0, 1]$ for which $f(x+)$, $f(x-)$ exist for each $x \in [0, 1]$ and

$$f(x) = \tfrac{1}{2}(f(x+) + f(x-)), \tag{16}$$

then Q may be metrized by defining for f, g, in Q

$$\sigma(f, g) = \sigma(G_f, G_g), \tag{17}$$

where G_f is a generalized graph of f:

$$G_f = \{(x, y) \mid x \in [0, 1]; f(x-) \leqslant y \leqslant f(x+) \text{ or } f(x+) \leqslant y \leqslant f(x-)\}.$$

It is not difficult to check that G_f is a closed set; so the definition (17) is justified. Also, for each x the set G_f determines uniquely the values $f(x+)$, $f(x-)$, and therefore, in view of (16), $f(x)$. Thus the correspondence between functions f in Q and their graphs is 1–1, and (17) defines a metric on Q.

If for Q we now take the set F_L of monotonically increasing functions defined on $[0, 1]$ which satisfy $f(0) = 0$ and $f(1) = L - 1$, and metrize F_L by (17), then associating f in F_L with its graph G_f gives an isometry between M_L and F_L, so from Theorem 4 we obtain:

COROLLARY 1. $H_\epsilon(F_L) \approx 1/\epsilon$.

In a similar way, if V_L is the set of functions f defined on $[0, 1]$ with $|f(x)| \leqslant M$ and

$$\operatorname*{Var}_{[0,1]} f \leqslant L - 1,$$

we obtain:

COROLLARY 2. $H_\epsilon(V_L) \approx 1/\epsilon$.

Proof. For $f \in V_L$, G_f can be represented parametrically as a curve with length not exceeding L, so associating f with G_f defines an isometry between V_L and a set which contains M_L and is contained in ${}_LX_2{}^*$. The conclusion then follows from Theorem 4.

As a final example, let Q_B be the set of functions defined on $[0, 1]$ for which $f(x+)$, $f(x-)$ exist, $|f(x)| \leqslant B$, and (16) is satisfied for $x \in [0, 1]$. Let Q_B be metrized by (17). Then:

THEOREM 5. $H_\epsilon(Q_B) \approx C_\epsilon(Q_B) \approx (1/\epsilon) \log (1/\epsilon)$.

Proof. To estimate $N_\epsilon(Q_B)$ from above, we construct an ϵ-net for Q_B in the following way: Take $n = n_\epsilon = [2/\epsilon] + 1 > 2/\epsilon$, and $x_i = i\delta$, $i = 0, 1, \ldots, n$, where $\delta = 1/n < \epsilon/2$. Let $f \in Q_B$ be given, and m_k, m_k', $k = 1, 2, \ldots, n$ be respectively the largest and smallest integers which satisfy

$$m_k\delta \leqslant \inf_{x\in[x_{k-1},x_k]} f(x) \leqslant \sup_{x\in[x_{k-1},x_k]} f(x) \leqslant m_k'\delta. \tag{18}$$

Let C_f consist of all points (x, y) which satisfy for some k, $k = 1, 2, \ldots, n$, the relation

$$x_{k-1} \leqslant x \leqslant x_k, \qquad m_k\delta \leqslant y \leqslant m_k'\,\delta.$$

We now show that

$$\sigma(G_f, C_f) \leqslant \epsilon. \tag{19}$$

Since C_f even contains G_f, $S_\epsilon(C_f) \supset G_f$. To show that $S_\epsilon(G_f) \supset C_f$, one checks that

$$\text{(20)}\quad \begin{cases} \text{if } (x_1, y_1),\ (x_2, y_2) \in G_f,\ x_1 < x_2, \text{ and } y_1 < c < y_2 \text{ or } y_2 < c < y_1, \text{ then} \\ \text{there exists } x_3 \text{ such that } x_1 \leqslant x_3 \leqslant x_2 \text{ and } (x_3, c) \in G_f. \end{cases}$$

https://doi.org/10.4153/CJM-1963-045-3 Published online by Cambridge University Press

From (20) it is seen that the $\sqrt{2}\,\delta$-neighbourhood of the part of G_f between the lines $x = x_{i-1}$ and $x = x_i$ contains all of the part of C_f between these lines, so $S_\epsilon(G_f) \supset C_f$, and (19) follows. Thus the collection $\{C_f\}_{f \epsilon Q_B}$ is an ϵ-net for Q_B. Each C_f is determined by the sequence $m_k, m_k', k = 1, 2, \ldots, n$. Since $|f(x)| \leqslant B$, $0 \leqslant x \leqslant 1$, each m_k, m_k' is one of $2[B/\delta] + 3$ integers m, $-[B/\delta] - 1 \leqslant m \leqslant [B/\delta] + 1$. Thus there are not more than $(2[B/\delta] + 3)^{2n}$ distinct sets in our net. Hence

$$N_\epsilon(Q_B) \leqslant (2[B/\delta] + 3)^{2n},$$
$$H_\epsilon(Q_B) \leqslant 2n \log (2[B/\delta] + 3) \prec (1/\epsilon) \log (1/\epsilon).$$

We obtain the estimate $C_\epsilon(Q_B) \succ (1/\epsilon) \log (1/\epsilon)$ by exhibiting a large number of ϵ-distinguishable functions. Take $n = n_\epsilon = [1/\epsilon] - 1$ and $\delta = 1/n > \epsilon$. For sequences of integers m_k, $k = 1, 2, \ldots, n$, $-[B/\delta] \leqslant m_k \leqslant [B/\delta]$, the corresponding step functions

$$s(0) = 0,\ s(x) = m_k\delta, \qquad x \in (x_{k-1}, x_k], \qquad k = 1, 2, \ldots, n,$$

are in Q_B, and are ϵ-distinguishable (in the Hausdorff metric). Since there are $(2[B/\delta] + 1)^n$ of these functions, we have

$$C_\epsilon(Q_B) \geqslant \log (2[B/\delta] + 1)^n \succ (1/\epsilon) \log (1/\epsilon),$$

which proves the theorem.

The author thanks his teacher, Professor G. G. Lorentz of Syracuse University, for suggesting these problems and for many helpful conversations concerning them.

References

1. F. Hausdorff, *Set theory* (New York, 1957).
2. A. N. Kolmogorov, *Asymptotic characteristics of some completely bounded metric spaces*, Dokl. Akad. Nauk. S.S.S.R., *108* (1956), 585–589 (Russian).
3. A. N. Kolmogorov and V. M. Tihomirov, *ϵ-entropy and ϵ-capacity of sets in function spaces*, Uspehi Mat. Nauk, *14* (1959), no. 2 (86), 3–86 (Russian).
4. ——— *ϵ-entropy and ϵ-capacity of sets in functional spaces*, Amer. Math. Soc. Translations, *17*, 2, 277–364. (English translation of **3.**)
5. G. G. Lorentz, *Metric entropies, widths and superpositions of functions*, Amer. Math. Monthly, *69* (1962), 469–485.
6. J. Riordan, *An introduction to combinatorial analysis* (New York, 1958).
7. A. G. Vituškin, *On the 13th problem of Hilbert*, Dokl. Akad. Nauk S.S.S.R., *95* (1954), 701–704 (Russian).
8. ——— *Estimation of the complexity of the tabulation problem* (Moscow, 1959, Russian).
9. ——— *Theory of the transmission and processing of information* (New York, 1961) (English translation of **8.**).

Syracuse University