ENTROPIES OF SEVERAL SETS OF REAL VALUED FUNCTIONS

G. F. CLEMENTS

Introduction. In this paper the entropies of several sets of real valued functions are calculated. The entropy of a metric set, a notion introduced by Kolmogorov [2], is a measure of its size in terms of the minimal number of sets of diameter not exceeding 2ε necessary to cover it. The most striking use of this notion to date has been given by Kolmogorov [4] and Vituškin [7] who 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 and other topics related to entropy see [5]. For other entropy calculations by the present author see [1]. The Kolmogorov-Vituškin result makes use of the following entropy calculation:

Let $F^n_{q=p+\alpha}(C, K) = F^n_q$ denote the class of real valued functions $f(x) = f(x_1, \cdots, x_n)$ defined on the unit cube S_n in the Euclidean n space which satisfy $|f(x)| \leqq C$ and have all partial derivatives of the order $k \leqq p$, with the pth order derivatives satisfying a Lipschitz condition of order α, $0 < \alpha \leqq 1$, with Lipschitz constant K:

$$|f^{(p)}(x) - f^{(p)}(x')| \leqq K|x - x'|^\alpha , \quad x, x' \in S_n .$$

Under the uniform metric ρ,

$$\rho(f, g) = \max_{x \in S_n} |f(x) - g(x)| ,$$

Kolmogorov [4, Th. XIV, p. 308] obtains

$$(1) \qquad H_\varepsilon(F^n_q) \asymp (1/\varepsilon)^{n/q} .$$

(The various symbols are defined below). In particular, with $p = 0$ and $n = 1$, this reads

$$(2) \qquad H_\varepsilon(\mathrm{Lip}_K\, \alpha) \asymp (1/\varepsilon)^{1/\alpha} ,$$

where we have written $\mathrm{Lip}_K\, \alpha$ in place of F^1_α.

The object of this paper is first to generalize (2) to sets of functions which satisfy a smoothness condition (§ 1), and second to show that

Received November 19, 1962. This research, part of the author's Ph. D. thesis, 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. AF 49 (638)-619.

(1) holds under the L_1 metric (§ 2).

Before stating our results more precisely we 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 $U \subseteqq W$ is called an ε-cover of A if for each U in γ, the diameter of $U, d(U)$, does not exceed 2ε, and $A \subseteqq \bigcup_{U \in \gamma} U$.

DEFINITION 2. A set $U \subseteqq W$ 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 $U \subseteqq W$ is said to be ε-distinguishable if the distance between any two points of U is greater than ε.

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

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

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

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

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

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

It is unusual to be able to determine these functions exactly and one is usually content with finding their order. We write $f(\varepsilon) \leqq g(\varepsilon)$ for $f(\varepsilon) = 0(g(\varepsilon))$ and $f(\varepsilon) \asymp g(\varepsilon)$ if both $f(\varepsilon) = 0(g(\varepsilon))$ and $g(\varepsilon) = 0(f(\varepsilon))$. Thus for various sets A we seek a function $h(\varepsilon)$ for which $H_\varepsilon \asymp h(\varepsilon)$ holds. The basic tool to this end is the following [4, Th. IV, p. 282]:

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

$$M_{2\varepsilon}(A) \leqq N_\varepsilon(A) \leqq N_\varepsilon^W(A) \leqq M_\varepsilon(A)$$

hold, and therefore

$$C_{2\varepsilon}(A) \leqq H_\varepsilon(A) \leqq C_\varepsilon(A) .$$

In § 1 we consider sets of continuous functions f defined on [0, 1] and satisfying a *smoothness condition.* The modulus of smoothness of f, $\omega^f(\varepsilon)$, is defined by

$$\omega^f(\varepsilon) = \max_{\substack{|t| \leqq \varepsilon \\ x \varepsilon [0\ 1]}} | \Delta_t f(x) | , \tag{3}$$

where $\Delta_t f(x)$ is the second difference of f at x with increment t:

$$\Delta_t f(x) = f(x + 2t) - 2f(x + t) + f(x) . \tag{4}$$

It is of course assumed that the maximum in (3) is taken over only those t for which (4) is defined.

For a fixed strictly increasing function $\phi(\varepsilon)$, let A_ϕ be the set of continuous functions f defined on $[0, 1]$ which satisfy $|f(x)| \leqq K$, and $\omega^f(\varepsilon) \leqq \phi(\varepsilon)$. With the uniform metric on A_ϕ, we give the best possible estimate from above for $H_\varepsilon(A_\phi)$ in the sense explained below (Th. 1). For the cases we examine, we will find the estimation of $H_\varepsilon(A_\phi)$ from below quite simple (Th. 2).

In § 2, we show that Kolmogorov's result (1) is also correct when the uniform metric is replaced by the L_1 metric ρ_{L_1} defined by:

$$\rho_{L_1}(f, g) = \int_{S_n} |f - g| \, dV = \int_0^1 \cdots \int_0^1 | f(x_1, \cdots, x_n) - g(x_1, \cdots, x_n) | \, dx_1 \cdots dx_n .$$

1. With A_ϕ as defined in the introduction, we now estimate $H_\varepsilon(A_\phi)$ from above:

THEOREM 1. *If* $\log(1/\phi(\varepsilon)) \leqq 1/\varepsilon$ *and* $M(\varepsilon) = \sum_{i=0}^{\infty} \phi(\varepsilon/2^i) < \infty$, *then*

$$H_\varepsilon(A_\phi) \leqq 1/M^{-1}(\varepsilon) .$$

This result is best possible within a constant factor; that is, there exists ϕ such that $H_\varepsilon(A_\phi) \geqq 1/M^{-1}(\varepsilon)$. In fact, with $\phi(\varepsilon) = \varepsilon$, one checks that $\mathrm{Lip}_{(1/2)}\, 1 \subset A_\phi$ and from Kolmogorov's result (2),

$$H_\varepsilon(\mathrm{Lip}_{(1/2)}\, 1) \geqq 1/\varepsilon = 1/M^{-1}(\varepsilon) .$$

The main idea in the proof of this theorem is that even though a function from A_ϕ may increase or decrease with arbitrary rapidity over a small interval, it will be approximated there well by its secant line. This is contained in the following lemma.

LEMMA 1. *Suppose* f *is defined and continuous on* $[x_0, x_0 + \delta]$ *and that* $\omega^f(\varepsilon) \leqq \phi(\varepsilon)$. *If* $F(\delta) = (1/2) \sum_{i=1}^{\infty} \phi(\delta/2^i) < \infty$ *and*

$$L(x) = f(x_0) + (x - x_0)(f(x_0 + \delta) - f(x_0))/\delta ,$$

then

$$|f(x) - L(x)| \leqq F(\delta) \quad \textit{for} \ x \in [x_0, x_0 + \delta] \, .$$

Proof. We shall show

$$(5) \qquad f(x) - L(x) \leqq F(\delta) \, , \quad x \in [x_0, x_0 + \delta] \, ;$$

the proof of

$$-F(\delta) \leqq f(x) - L(x) \, , \quad x \in [x_0, x_0 + \delta]$$

is similar. To prove (5), it is sufficient to prove the inequalities

$$(6) \qquad f(x_k^i) - L(x_k^i) \leqq (1/2) \sum_{i=1}^{k} \phi(\delta/2^i) \, , \quad i = 0, 1, \cdots, 2^k \, ; \quad k = 1, 2, \cdots$$

where $x_k^i = x_0 + (i/2^k)\delta$. For if (6) is established (5) follows from the continuity of $f(x) - L(x)$.

We prove (6) by induction. For $k = 1$, we have

$$\begin{aligned} &f(x_0 + \delta) - 2f(x_0 + \delta/2) + f(x_0) = \Delta_{\delta/2} f(x_0) \geqq -\phi(\delta/2) \, , \\ &f(x_0 + \delta/2) \leqq \{f(x_0) + (f(x_0 + \delta) - f(x_0))/2\} + (1/2)\phi(\delta/2) \\ &\qquad = L(x_1^1) + (1/2)\phi(\delta/2) \, , \end{aligned}$$

and

$$f(x_1^1) - L(x_1^1) \leqq (1/2)\phi(\delta/2) \, .$$

We also have $f(x_1^0) - L(x_1^0) = f(x_1^2) - L(x_1^2) = 0$, so (6) is established for $k = 1$. Assuming the inequalities (6) hold for k, we consider them for $k + 1$. Let i be given, $0 \leqq i \leqq 2^{k+1}$. If i is even,

$$x_{k+1}^i = x_0 + (i/2^{k+1})\delta = x_0 + (i/2)(1/2^k)\delta = x_k^{i/2}$$

and

$$f(x_{k+1}^i) - L(x_{k+1}^i) \leqq (1/2) \sum_{i=1}^{k} \phi(\delta/2^i) \leqq (1/2) \sum_{i=1}^{k+1} \phi(\delta/2^i)$$

by the induction hypothesis. If i is odd, we have

$$x_{k+1}^i = x_k^{(i-1)/2} + \delta/2^{k+1} \, ; \quad x_{k+1}^i = x_k^{(i+1)/2} - \delta/2^{k+1}$$

and

$$f(x_k^{(i+1)/2}) - 2f(x_{k+1}^i) + f(x_k^{(i-1)/2}) = \Delta_{\delta/2^{k+1}} f(x_{k+1}^{(i-1)/2}) \geqq -\phi(\delta/2^{k+1})$$

or

$$(7) \qquad f(x_{k+1}^i) \leqq (1/2)\{f(x_k^{(i-1)/2}) + f(x_k^{(i+1)/2})\} + (1/2)\phi(\delta/2^{k+1}) \, .$$

By the induction hypothesis, $f(x_k^{(i-1)/2})-L(x_k^{(i-1)/2})$ and $f(x_k^{(i+1)/2})-L(x_k^{(i+1)/2})$ do not exceed $(1/2)\sum_{i=1}^{k}\phi(\delta/2^i)$, so from (7) we have

$$\begin{aligned} f(x_{k+1}^i) &\leqq (1/2)\{f(x_k^{(i-1)/2}) - L(x_k^{(i-1)/2}) + f(x_k^{(i+1)/2}) - L(x_k^{(i+1)/2})\} \\ &\quad + (1/2)\{L(x_k^{(i-1)/2}) + L(x_k^{(i+1)/2})\} + (1/2)\phi(\delta/2^{k+1}) \\ &\leqq (1/2)\sum_{i=1}^{k}\phi(\delta/2^i) + L(x_{k+1}^i) + (1/2)\phi(\delta/2^{k+1}) , \end{aligned}$$

and

$$f(x_{k+1}^i) - L(x_{k+1}^i) \leqq (1/2)\sum_{i=1}^{k+1}\phi(\delta/2^i) .$$

Thus (6) and the lemma follow.

Proof of the theorem. Let $\varepsilon > 0$ be given. Put $n = n_\varepsilon = [1/\varepsilon] + 1$ (here and below $[x]$ denotes the largest integer not exceeding x), $\delta = 1/n < \varepsilon$, and $\eta = \phi(\delta)$. With $f \in A_\phi$, associate the sequence

$$S_f\colon k_1, k_2, \cdots, k_n \tag{8}$$

where $k_i = [f(i\delta)/\eta]$, $i = 1, 2, \cdots, n$. Notice that for given k_i and k_{i+1}, k_{i+2} takes on one of seven values. This is because

$$\gamma = f((i+2)\delta) - 2f((i+1)\delta) + f(i\delta)$$

is a second difference with increment δ, so

$$-\phi(\delta) \leqq \gamma \leqq (k_{i+2} + 1 - 2k_{i+1} + k_i + 1)\eta \tag{9}$$

and

$$(k_{i+2} - 2k_{i+1} - 2 + k_i)\eta \leqq \gamma \leqq \phi(\delta) . \tag{10}$$

From (9) and (10) we have

$$-1 = -\phi(\delta)/\eta \leqq k_{i+2} + \{2 - 2k_{i+1} + k_i\} \leqq \phi(\delta)/\eta + 4 = 5 ;$$

hence if $q = -\{2 - 2k_{i+1} + k_i\}$, k_{i+2} is one of

$$q - 1, q, q + 1, \cdots, q + 5 .$$

Since $|f(x)| \leqq K$ for $x \in [0, 1]$, k_1 and k_2 are each one of $2[K/\eta] + 1 < 3[K/\eta]$ integers. Then the number of distinct sequences S_f does not exceed

$$(3K/\eta)^2 7^n .$$

With S_f we associate the function P_f, the graph of which is the polygonal line determined by the points $(i\delta, k_i\eta)$, $1 = 1, 2, \cdots, n$. For $x \in [0, 1]$, it follows from the lemma that

$$|f(x) - P_f(x)| \leqq |f(x) - L(x)| + |L(x) - P_f(x)|$$
$$\leqq F(\delta) + \eta = F(\delta) + \phi(\delta) \leqq F(\varepsilon) + \phi(\varepsilon) = M(\varepsilon) .$$

Thus $\rho(f, P_f) \leqq M(\varepsilon)$ and the set $\{P_f\}_{f\in A_\phi}$ is an $M(\varepsilon)$-net for A_φ. Since the functions P_f and the sequences S_f correspond in a one-to-one way and $\delta \asymp \varepsilon$, we have

$$N_{M(\varepsilon)}(A_\phi) \leqq (3K/\eta)^2 7^n ,$$

and

$$H_{M(\varepsilon)}(A_\phi) \leqq \log(1/\eta) + 1/\varepsilon \leqq \log(1/\phi(\varepsilon)) + 1/\varepsilon \preceq 1/\varepsilon ;$$

hence,

$$H_\varepsilon(A_\phi) \preceq 1/M^{-1}(\varepsilon) .$$

When $\phi(\varepsilon)$ is concave and strictly increasing, $H_\varepsilon(A_\phi)$ may be estimated from below in the following way. Take

$$n = [1/\phi^{-1}(\varepsilon)] - 1 , \quad \delta = 1/n > \phi^{-1}(\varepsilon) , \quad x_i = i\delta , \qquad i = 1, 2, \cdots, n .$$

With each sequence of positive and negative ones

$$m_1, m_2, \cdots, m_{[n/2]} \tag{11}$$

associate the function $f = f_{m_1, m_2, \cdots, m_{n/2}}$ defined by

$$\begin{aligned} f(x) &= (1/2)m_i\phi(x - x_{2i-2}) , & x &\in [x_{2i-2}, x_{2i-1}) \\ f(x) &= -(1/2)m_i\{\phi(x - x_{2i-1}) - \phi(\delta)\} & x &\in [x_{2i-1}, x_{2i}) \\ f(x) &= 0 & x &\in [x_{2[n/2]}, 1] , \end{aligned}$$

$i = 1, 2, \cdots, [n/2]$.

Each of these functions is in A_ϕ since $|f(x + \varepsilon) - f(x)| \leqq (1/2)\phi(\varepsilon)$ by the concavity of $\phi(\varepsilon)$, and the set D of all such functions is ε-distinguishable since each pair of functions in D differ by $2(1/2)\phi(\delta) > \varepsilon$ at some x_i. Since there are $2^{[n/2]}$ sequences (11) and therefore $2^{[n/2]}$ functions in D, we have

$$M_\varepsilon(A_\phi) \geqq 2^{[n/2]} , \quad \text{and} \quad C_\varepsilon(A_\phi) \succeq n \succeq 1/\phi^{-1}(\varepsilon) .$$

This proves:

THEOREM 2. *If $\phi(\varepsilon)$ is concave and strictly increasing, then*

$$C_\varepsilon(A_\phi) \succeq 1/\phi^{-1}(\varepsilon) .$$

EXAMPLES. If $\phi_\alpha(\varepsilon) = \varepsilon^\alpha$, $0 < \alpha \leqq 1$, then for the class A_{ϕ_α} we have $M(\varepsilon) = \sum_{i=0}^{\infty} (\varepsilon/2^i)^\alpha \asymp \varepsilon^\alpha$, so Theorems 1 and 2 give

(12) $$C_\varepsilon(A_{\phi_\alpha}) \asymp H_\varepsilon(A_{\phi_\alpha}) \asymp 1/M^{-1}(\varepsilon) = 1/\varepsilon^{1/\alpha} .$$

For $0 < \alpha < 1$, it is known that

(13) $$\mathrm{Lip}_{K_1^\alpha} \subset A_{\phi_\alpha} \subset \mathrm{Lip}_{K_2^\alpha}$$

for suitable K_1, K_2. Since the entropy of Lip_{K^α} is independent of K [4, p. 286], this inclusion and (12) give (2). If $\alpha = 1$, $\mathrm{Lip}_K 1$ is properly contained in A_{ϕ_α}. For example, the function

$$f(x) = \begin{cases} (1/2 \log 2)x \log x & x \neq 0 \\ 0 & x = 0 \end{cases}$$

is not in $\mathrm{Lip}_K 1$ since $f'(x)$ is unbounded on $(0, 1]$. But f is in A_{ϕ_1} since one may verify that $|\Delta_t f(x)| \leqq |t|$ for $x \in [0, 1]$ and therefore $\omega^f(h) \leqq h$.

Also, if $\phi(\varepsilon) = \varepsilon + \varepsilon \log (1/\varepsilon)$, our results give for A_ϕ, which is intermediate between Lip 1 and each of the classes Lip α, $0 < \alpha < 1$, the estimate $H_\varepsilon(A_\phi) \asymp (1/\varepsilon) \log (1/\varepsilon)$.

Our Theorems 1 and 2 thus contain the special case (2) of (1) and somewhat more.

2. We now show that (1) also holds under the L_1 metric.

THEOREM 3. *Under the L_1 metric, $H_\varepsilon(F_q^n) \asymp (1/\varepsilon)^{n/q}$.*

Proof. Since the L_1 metric is smaller than the uniform metric, the estimate $H_\varepsilon(F_q^n) \lesssim (1/\varepsilon)^{n/q}$ is immediate from Kolmogorov's result (1). To get the reverse estimate we show the existence of a large number of ε/M-distinguishable functions in F_q^n without actually producing them. M is a constant which will be implicity determined later.

The functions we seek are among those given by Kolmogorov [4, p. 311] to establish the estimate $H_\varepsilon(F_q^n) \gtrsim (1/\varepsilon)^{n/q}$ in the uniform metric. Set

$$\phi(y) = \phi((y_1, y_2, \cdots, y_n)) = \begin{cases} a\Pi_{i=1}^n(1 + y_i)^q , & |y_i| \leqq 1 , \quad i = 1, 2, \cdots, n \\ 0 & \text{otherwise} . \end{cases}$$

Put $\Delta = (\varepsilon/a)^{1/q}$ and let $x^0, x^1, \cdots, x^s$ be a maximal 2Δ-distinguishable set in S_n. It is clear that

$$s \asymp 1/\Delta^n \asymp 1/\varepsilon^{n/q} .$$

Let U consist of all functions of the form

$$f(x) = f_{j_1, j_2, \cdots, j_s}(x) = \sum_{r=0}^{s} j_r \Delta^q \phi\left(\frac{x - x^r}{\Delta}\right)$$

$$= \sum_{r=0}^{s} j_r \varepsilon \Pi_{i=1}^{n}\left(1 + \frac{x_i - x_i^r}{\Delta}\right)^q\left(1 - \frac{x_i - x_i^r}{\Delta}\right)^q ,$$

where $j_r = \pm 1$, $r = 0, 1, \cdots, s$.

For suitable a and small ε, U is contained in F_q^n. U is ε-distinguishable in the uniform metric, but not in the L_1 metric; however we can show the existence of a subset of U which is ε/M-distinguishable in the L_1 metric and contains enough functions for our purpose. We do this as follows: Let $k(\varepsilon)$ be the largest integer such that for each function f of U there exist no more than $k(\varepsilon)$ other functions f' of U which satisfy

$$\rho_{L_1}(f, f') \leqq \varepsilon/M . \tag{14}$$

If one now selects $f^{(1)}$ arbitrarily from U and with it all functions of $U f_1^{(1)}, \cdots, f_{r(1)}^{(1)}$, $r(1) \leqq k(\varepsilon)$, which satisfy (14) with $f = f^{(1)}$, and then from the remaining functions of U selects $f^{(2)}$ arbitrarily and with it all functions of $U f_1^{(2)}, \cdots, f_{r(2)}^{(2)}$, $r(2) \leqq k(\varepsilon)$, which satisfy (14) with $f = f^{(2)}$, and so on until U is exhausted, one obtains at least $t = [(2^s)/(k(\varepsilon) + 1)]$ groups of functions. The functions $f^{(1)}, f^{(2)}, \cdots, f^{(t)}$ are mutually more than ε/M apart in the L_1 metric and therefore form an ε/M-distinguishable subset of F_q^n. Thus

$$M_{\varepsilon/M}(F_q^n) \geqq \frac{2^s}{k(\varepsilon) + 1}$$

and

$$H_{\varepsilon/2M}(F_q^n) \geqq s - \log_2 (k(\varepsilon) + 1) \geqq c(1/\varepsilon)^{n/q} - \log_2(k(\varepsilon) + 1) \tag{15}$$

where $c > 0$. We will show that when M is taken favorably

$$\log_2 (k(\varepsilon) + 1) \leqq (1/2)c(1/\varepsilon)^{n/q} ,$$

so from (15) it will follow that

$$H_{\varepsilon/2M}(F_q^n) \geqq (1/\varepsilon)^{n/q} , \quad \text{or} \quad H_\varepsilon(F_q^n) \geqq (1/\varepsilon)^{n/q} ,$$

completing the proof.

To estimate $k(\varepsilon)$ from above notice that for functions $f_1 = f_{j_1^1, \cdots, j_s^1}$ and $f_2 = f_{j_1^2, \cdots, j_s^2}$ of U, the inequality

$$\rho_{L_1}(f_1, f_2) \leqq \varepsilon/M \tag{16}$$

implies that $j_i^1 \doteq j_i^2$ with at most $c_1/(K^{n/q})$ exceptions, for some constant c_1. This is because if $j_m^1 \neq j_m^2$, then

$$\int_{G_\Delta(x^m)} |f - g|\, dV = 2\varepsilon \Delta^n \left\{\int_{-1}^{1} (1 - t^2)^q dt\right\}^n = (1/c_1)\varepsilon^{(n/q)+1} ,$$

where

$$G_\Delta(x^m) = G_\Delta(x_1^m, \cdots, x_n^m)$$
$$= \{(x_1, x_2, \cdots, x_n) \mid |x_i - x_i^m| \leqq \Delta, i = 1, 2, \cdots, n\} .$$

Thus for fixed $f_1 \in U$, the number of functions $f_2 \in U$ which satisfy (16) does not exceed

$$\sum_{i=0}^{[c_1/M\varepsilon^{n/q}]} \binom{s}{i} \leqq ([c_1/M\varepsilon^{n/q}] + 1)\binom{[c_2/\varepsilon^{n/q}]}{[c_1/M^{n/q}]} ;$$

therefore

$$\log_2 k(\varepsilon) \leqq \log_2 ([c_1/M\varepsilon^{n/q}] + 1) + \log_2 \binom{[c_2/\varepsilon^{n/q}]}{[c_1/M\varepsilon^{n/q}]} .$$

If M is taken suitably large, one finds from Stirling's formula that

$$\log_2 \binom{n}{n/M} \leqq (1/4)cn$$

for large n. Then for small ε and a suitable M, we have

$$\log_2 k(\varepsilon) \leqq (1/2)c(1/\varepsilon)^{n/q}$$

and the theorem follows.

Since functions of the class Lip 1 are functions of bounded variation, the above calculation accomplishes part of showing that $H_\varepsilon(V) \asymp 1/\varepsilon$, in the L_1 metric where V is the set of functions f defined but not necessarily continuous on $[0, 1]$, which satisfy $|f(x)| \leqq M$, and $\mathrm{Var}_{[0\,1]} f \leqq B$, where B is a constant not depending on f.

COROLLARY. $H_\varepsilon(V) \asymp 1/\varepsilon$.

Proof. Since $V \supset \mathrm{Lip}\, 1$, $H_\varepsilon(V) \geqq 1/\varepsilon$ follows from the theorem. To get the reverse estimate, take $n = [1/\varepsilon]$, $\delta = 1/n$ and $x_i = i\delta/2$, $i = 0, 1, \cdots, n$. For given $f \in V$, let m_{2i-2} be the largest integer such that $\delta m_{2i-2} < f(x)$ for all $x \in [x_{2i-2}, x_{2i}]$ and let m_{2i-1} be the smallest integer such that $\delta m_{2i-1} > f(x)$ for all $x \in [x_{2i-2}, x_{2i}]$, $i = 1, 2, \cdots, n$. If $g_f(x)$ is the function the graph of which is the polygonal line determined by the points

$$(x_i, \delta m_i),\ i = 0, 1, \cdots, (2n - 1) , \quad \text{and} \quad (1, \delta m_{2n-1}) ,$$

we claim that $\rho_{L_1}(f, g_f) \leqq c\varepsilon$ for $\varepsilon < (1/2)$, where $c = 2(L + 2)$ is independent of ε. This is because

$$|f(x) - g(x)| \leqq \delta m_{2i-1} - \delta m_{2i-2}\,, \quad x \in [x_{2i-2}, x_{2i}]\,,$$

and therefore

$$\int_{x_{2i-2}}^{x_{2i}} |f(x) - g(x)|\,dx \leqq (m_{2i-1} - m_{2i-2})\delta^2\,.$$

Then

$$\rho_{L_1}(f, g) = \int_0^1 |f(x) - g(x)|\,dx \leqq \delta^2 \sum_{i=1}^{n} (m_{2i-1} - m_{2i-2})\,.$$

It is clear that

$$\delta \sum_{i=1}^{n} (m_{2i-1} - 1 - (m_{2i-2} + 1)) \leqq \mathrm{Var}_{[0,1]} f \leqq B\,,$$

so

$$\rho_{L_1}(f, g) \leqq \delta(B + 2n\delta) \leqq c\varepsilon\,.$$

Thus the functions $\{g_f(x)\}_{f \in V}$ form a $c\varepsilon$-net for V. We now estimate from above how many functions are in this net. To do this notice that labeling the function $g_f(x)$ with the (finite) sequence

$$n_0, n_1, \cdots, n_{2n-1} \tag{17}$$

where $n_0 = m_0$ and

$$n_i = (-1)^{i+1}(m_i - m_{i-1}) \geqq 0\,, \qquad i = 1, 2, \cdots, 2n - 1$$

gives a one-to-one correspondence between the functions in our net and some sequences of the form (17). It therefore suffices to estimate how many different sequences (17) will be required to label the functions in the net. Since $(n_i - 2)\delta \leqq \mathrm{Var}_{[x_{i-1}, x_i]} f$, we have

$$\delta n_0 + \delta \sum_{i=1}^{2n-1} (n_i - 2) \leqq \delta M/\delta + B\,,$$

or

$$\sum_{i=0}^{2n-1} n_i \leqq M/\delta + B/\delta + 4/\delta = B'/\delta\,,$$

so the nonzero terms among (17) form a composition [6] of not more than $2n$ parts of an integer $k \leqq B'/\delta$. Since the number of compositions of k with exactly i parts is $\binom{k-1}{i-1}$ [6, p. 124], and $2n - i$ zeros can fall in $\binom{2n}{2n-i}$ ways among $2n$ places, the number of labels (17) with i nonzero parts which add to k does not exceed

$$\binom{2n}{2n-1}\binom{k-1}{i-1}.$$

Then in all there are not more than

$$\sum_{k=0}^{B'/\delta}\sum_{i=1}^{\min(2n,k)}\binom{2n}{2n-i}\binom{k-1}{i-1} \leqq (B'/\delta)2n\binom{2n}{n}\binom{B'/\delta}{B'/2\delta}$$

functions in our $c\varepsilon$-net; hence

$$N_{c\varepsilon}(V) \leqq B'2n^2\binom{2n}{n}\binom{B'n}{B'n/2}.$$

Since $\log\binom{n}{n/2} \leqq n$, we finally obtain

$$H_{c\varepsilon}(V) \leqq \log n + \log\binom{2n}{n} + \log\binom{B'n}{B'n/2} \leqq 2n + B'n \leqq n \leqq 1/\varepsilon ,$$

or

$$H_\varepsilon(V) \leqq 1/\varepsilon .$$

The author is indebted to his teacher, Prof. G. G. Lorentz of Syracuse University, for suggesting these problems and for many helpful conversations concerning them.

References

1. G. F. Clements, *Entropies of sets of functions of bounded variation*, Canad. J. Math., **15** (1963), 422-432.
2. A. N. Kolmogorov, *Asymptotic characteristics of some completely bounded metric spaces*, DAN, **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. Translation, 17, 2, 277-364 (English translation of 3).
5. G. G. Lorentz, *Metric entropy, 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*, DAN, **95** (1954), 701-704 (Russian).

UNIVERSITY OF COLORADO