EPSILON ENTROPY OF STOCHASTIC PROCESSES[1]

BY EDWARD C. POSNER, EUGENE R. RODEMICH, AND HOWARD RUMSEY, JR.

Jet Propulsion Laboratory, California Institute of Technology

0. Summary. This paper introduces the concept of *epsilon-delta entropy* for "probabilistic metric spaces." The concept arises in the study of efficient data transmission, in other words, in "Data Compression." In a case of particular interest, the space is the space of paths of a stochastic process, for example $L_2[0, 1]$ under the probability distribution induced by a mean-continuous process on the unit interval.

For any epsilon and delta both greater than zero, the epsilon-delta entropy of any probabilistic metric space is finite. However, when delta is zero, the resulting entropy, called simply the *epsilon entropy* of the space, can be infinite. We give a simple condition on the eigenvalues of a process on $L_2[0, 1]$ such that any process satisfying that condition has finite epsilon entropy for any epsilon greater than zero. And, for any set of eigenvalues not satisfying the given condition, we produce a mean-continuous process on the unit interval having infinite epsilon entropy for every epsilon greater than zero. The condition is merely that $\sum n\sigma_n^2$ be finite, where $\sigma_1^2 \geqq \sigma_2^2 \geqq \cdots$ are the eigenvalues of the process.

1. Introduction. This subject is motivated by the following considerations relating to efficient data transmission, a subject that has come to be known as "Data Compression." Suppose one has an experimental source given by some probability law which is assumed known. One wishes to transmit the outcomes to a remote place using as few "bits" of information as possible. There is assumed to be a certain "fidelity criterion" such that the actual outcome occurring is to be known to the recipient of the message within a given fidelity after he receives the message. One is also allowed to ignore a certain fraction of the outcomes, that is, one does not attempt to transmit them. What is the minimum number of bits that one can get away with under these constraints? That is, what is the best way to "compress" the data?

We idealize this situation by defining the concept of *probabilistic metric space* and its $\epsilon;\delta$ *entropy* $H_{\epsilon;\delta}(X)$. Thus, define a probabilistic metric space $X = (X, d, \mu)$ as follows:

(1) (X, d) is to be a complete separable metric space, of points of the set X under the metric d.

(2) (X, μ) is to be a probability space, where the Borel field of sets on which μ is defined is to be the completion of the class of Borel sets in (X, d), that is, the completion of the Borel field generated by the open sets of X.

The probability space (X, μ) represents the space of experimental outcomes.

Received 16 December 1966.

[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.

The metric d represents the fidelity criterion, stating how close one outcome is to another.

Now any transmission system, it is reasonable to assume, works as follows: Given the received message, there is a certain Borel set of outcomes of X that could have arisen and still yielded the same message. The requirement that outcomes be known within ϵ can be translated as the requirement that the Borel sets occurring in this way have diameters at most ϵ. Thus, we are led to consider partitions of X by measurable sets of diameters at most ϵ (called briefly *ϵ-sets*). However, since we are allowed to ignore a certain fraction, say δ at most, of X in designing our data compression system, we need only require that the union of the sets in the ϵ-partition of X has probability at least $1 - \delta$. Such a partition we call an *ϵ; δ partition*, which is defined for $\epsilon < 0$, $\delta \geqq 0$ (when $\delta = 0$, we call such partitions briefly *ϵ-partitions* of X). We consider only partitions by a finite or denumerably infinite number of sets, or equivalently, partitions in which all but a denumerable number of the sets lie in a set of probability 0.

Now when we have an $\epsilon;\delta$ partition of X, the number of bits necessary to describe into which set of the partition the outcome actually falls is given by a well-known formula of Shannon [9]. Namely, let p_i denote the probability of the ith set U_i of the partition U which has positive probability. Let $\{q_i\}$ denote the discrete distribution $\{p_i/\sum p_i\}$. This distribution has an entropy; we call this entropy $H(U)$, the entropy of the partition U (given in logarithms to base e):

$$H(U) = \sum q_i \log q_i^{-1}. \tag{1}$$

Thus, $H(U)$ is the minimum number of bits (except for a factor log 2 which we ignore here and elsewhere) necessary to describe into which set of U the outcome falls, given that it falls into the part of X covered by some set of the ϵ; δ partition. However, our ϵ; δ partition may not have been especially well chosen. Thus, it is natural to consider for the class $\mathfrak{U}_{\epsilon;\delta}$ of $\epsilon;\delta$ partitions of the probabilistic metric space X the quantity $H_{\epsilon;\delta}(X)$ defined by

$$H_{\epsilon;\delta}(X) = \inf_{U \epsilon \mathfrak{U}_{\epsilon;\delta}} H(U). \tag{2}$$

This quantity $H_{\epsilon;\delta}(X)$ is called the *$\epsilon;\delta$ entropy of X* (when $\delta = 0$, we write $H_\epsilon(X)$ and call it the *ϵ entropy of X*). It will later be proved that the infimum is actually achieved for some ϵ; δ partition U in $\mathfrak{U}_{\epsilon;\delta}$.

The ϵ; δ entropy of X then measures the smallest number of bits necessary to describe at least $1 - \delta$ of X by measurable sets of diameter at most ϵ. A further discussion of the exact relevance of this concept to Data Compression is found in [5].

The way $H_{\epsilon;\delta}(X)$ is defined leaves open the possibility that $H_{\epsilon;\delta}(X)$ is infinite. That is, no ϵ; δ partition might have finite entropy. However, when $\delta > 0$, we can prove that $H_{\epsilon;\delta}(X)$ is finite. To prove this, we produce a finite ϵ; δ partition, which of course has finite entropy.

First observe that ϵ; δ partitions always exist (in fact, ϵ-partitions always exist). For let $\{x_i\}$ be a countable dense subset of X, and let U_i denote the sphere of radius $\epsilon/2$ about x_i. Then the U_i cover X. Let $V_1 = U_1$, $V_2 = U_2 - U_1$, $V_3 =$

$U_3 - (U_1 \cup U_2), \cdots, V_n = U_n - \bigcup_{i=1}^{n-1} U_i, \cdots$. Then each V_i is measurable and has diameter at most ϵ. Furthermore, $\bigcup_{i=1}^{\infty} V_i = X$, and the V_i are disjoint. Hence, V_i is indeed an ϵ-partition of X.

Now let $\{V_i\}$ be any ϵ-partition of X, and let $p_i = \mu(V_i)$. Thus $\sum_{i=1}^{\infty} p_i = 1$. Let i_0 be such that $\sum_{i>i_0} p_i \leqq \delta$. Then $\{V_i, 1 \leqq i \leqq i_0\}$ is a finite $\epsilon; \delta$ partition of X. Hence, $H_{\epsilon;\delta}(X)$ is finite if $\delta > 0$.

However, it turns out that $H_\epsilon(X)$ can be infinite. It is of special interest in this subject to give conditions on certain classes of probabilistic metric spaces which makes the ϵ-entropy finite. In this paper, we shall do this for the class of probabilistic metric spaces given by $L_2[0, 1]$ under the measure induced by a mean-continuous stochastic process. Before we do this, however, we shall give some results on the continuity of $H_{\epsilon;\delta}(X)$ in ϵ and δ valid for arbitrary probabilistic metric spaces.

2. Semicontinuity of epsilon-delta entropy. It is the purpose of this section to prove a lower semicontinuity result about the function $H_{\epsilon;\delta}(X)$. Two preliminary lemmas are needed, and will now be proved.

LEMMA 1. *For $n = 1, 2, \cdots$, let U_n be a partition of part of a probability space with sets of probabilities $\{p_k^{(n)}\}$ arranged in non-increasing order, $\sum_{k=1}^{\infty} p_k^{(n)} = \mu_n$. Let*

$$\lim_{n\to\infty} p_k^{(n)} = p_k, \qquad k = 1, 2, \cdots,$$

$$\lim_{n\to\infty} \mu_n = \mu > 0.$$

If the $\{U_n\}$ have entropies bounded by a constant B, then $\sum p_k = \mu$, and

$$\sum (p_k/\mu) \log \mu/p_k \leqq B.$$

PROOF. For any k_1,

$$\sum_{k=1}^{k_1} p_k = \lim_{n\to\infty} \sum_{k=1}^{k_1} p_k^{(n)} \leqq \lim_{n\to\infty} \mu_n = \mu,$$

so $\sum_{k=1}^{\infty} p_k \leqq \mu$. Choose $\beta > 0$, and k_2 such that $p_k < \beta$ for $k > k_2$. Then for n sufficiently large, $p_k^{(n)} < \beta$ for $k > k_2$. After the first k_2 sets of U_n, the remaining, if any, must all have measure less than β. Their total measure is $\mu_n - \sum_{k=1}^{k_2} p_k^{(n)}$. Hence

$$B \geqq H(U_n) \geqq [(\mu_n - \sum_{k=1}^{k_2} p_k^{(n)})/\beta]\cdot(\beta/\mu_n) \log (\mu_n/\beta);$$

therefore,

$$\sum_{k=1}^{k_2} p_k^{(n)} \geqq \mu_n - [(1/\mu_n) \log (\mu_n/\beta)]^{-1} B.$$

Let $n \to \infty$; then $k_2 \to \infty$; then $\beta \to 0$. The result is $\sum_{k=1}^{\infty} p_k \geqq \mu$. Hence this sum equals μ.

The inequality to be proved follows from the inequality for finite sums:

$$\sum_{k=1}^{k_2} (p_k/\mu) \log (\mu/p_k) = \lim_{n\to\infty} \sum_{k=1}^{k_2} (p_k^{(n)}/\mu_n) \log (\mu_n/p_k^{(n)})$$
$$\leqq \limsup_{n\to\infty} H(U_n) \leqq B.$$

Lemma 1 is proved.

LEMMA 2. *Let* $\{p_1, p_2, \cdots\}$ *and* $\{q_1, q_2, \cdots\}$ *be two sequences of non-negative numbers such that* $p_1 \geqq p_2 \geqq \cdots$,

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} q_k \leqq 1,$$

and for any positive integer n,

$$\sum_{k=1}^{n} p_k \leqq \sum_{k=1}^{n} q_k .$$

Then

$$\sum_{k=1}^{\infty} q_k \log (1/q_k) \leqq \sum_{k=1}^{\infty} p_k \log (1/p_k). \tag{3}$$

PROOF. Suppose first that both sequences are zero after some integer N. The proof in this case depends on the fact that the $\{q_k\}$ can be transformed into the $\{p_k\}$ in a finite number of steps, changing only two at a time, always increasing the sum on the left in (3). Let there be m values of j for which $q_j \neq p_j$. The proof is by induction on m, $0 \leqq m \leqq N$. Equality holds when $m = 0$. Suppose $m = M > 0$, and the result is known true for $m < M$.

Since $\sum q_j = \sum p_j$, there is a first index r for which $q_r > p_r$ and a first index s for which $q_s < p_s$. By hypothesis, $r < s$, $q_r > p_r \geqq p_s > q_s$. If we replace q_r and q_s by values q_r', q_s' such that

$$q_r' + q_s' = q_r + q_s , \qquad q_r > q_r' \geqq q_s' > q_s ,$$

the value of $\sum q_j \log (1/q_j)$ is increased, since the function $x \log (1/x)$ has a negative second derivative. This may be done so that either

$$q_r' = p_r , \qquad p_s \geqq q_s', \quad \text{or} \quad q_r' \geqq p_r , \qquad p_s = q_s'.$$

It is easily verified that the new set of q_j satisfy the hypotheses. By the induction hypothesis, the inequality (3) is true. Lemma 3 is proved for finite sequences.

Now we consider the general case. Let K be any positive integer, and define

$$p^* = \sum_{k=K+1}^{\infty} p_k , \qquad q^* = \sum_{k=K+1}^{\infty} q_k .$$

The sequences $\{p_1, \cdots, p_K, p^*, 0, 0, \cdots\}$ and $\{q_1, \cdots, p_K, q^*, 0, 0, \cdots\}$ satisfy the hypotheses. Hence by the above,

$$\sum_{k=1}^{K} q_k \log (1/q_k) + q^* \log (1/q^*) \leqq \sum_{k=1}^{K} p_k \log (1/p_k) + p^* \log (1/p^*).$$

The two sides of this inequality approach the members of (3) as $K \to \infty$. Hence (3) is true.

This completes the proof of Lemma 2, and we come to Theorem 1.

THEOREM 1. *In a probabilistic metric space* (X, d, μ), *let* $\{U_n\}$ *be a sequence of* ϵ_n ; δ_n*-partitions, with* $\epsilon_n \to \epsilon$, $\delta_n \to \delta$ *as* $n \to \infty$, *with* $\epsilon > 0$, $\delta \geqq 0$. *Then there is an* ϵ; δ*-partition* U *with*

$$H(U) \leqq \liminf_{n\to\infty} H(U_n).$$

PROOF. The proof uses the fact that the closed subsets of a compact metric space themselves form a compact metric space under a suitable metric; this compactness enables one to construct a limit covering from the $\{U_n\}$. Of course, X

need not be compact, so this limit is first obtained in compact sets approximating X in measure. The possibility of the existence of atoms in X (points of positive probability) causes additional technical complications.

We can assume

$$\liminf_{n\to\infty} H(U_n) = L = \lim_{n\to\infty} H(U_n),$$

by taking a subsequence. Let $U_n = \{A_{rk}\}$, with $\mu(A_{nk}) = p_{nk}$, $p_{n1} \geqq p_{n2} \geqq \cdots$, and define

$$m_n = \sum_{k=1}^{\infty} p_{nk} .$$

Then $m_n \geqq 1 - \delta_n$. We can take a subsequence of $\{U_n\}$ by the diagonal process, such that the limits $p_k = \lim_{n\to\infty} p_{rk}$ exist for all k, as well as $m = \lim_{n\to\infty} m_n$. By Lemma 1, $\sum_{k=1}^{\infty} p_k = m \geqq 1 - \delta$, and

$$\sum_{k=1}^{\infty} (p_k/m) \log (m/p_k) \leqq L. \tag{4}$$

The partition U will be an ϵ-partition covering a set of measure m.

If X contains atoms, we need to take another subsequence to control the way the $\{U_r\}$ cover the atoms. For any given atom x_j, there is a k_1 such that $p_{k_1} < \mu(x_j)$. For n sufficiently large, $p_{n,k_1} < \mu(x_j)$. Then x_j cannot lie in a set of U_n except the first $k_1 - 1$. Hence there is a subsequence of $\{U_n\}$ on which either x_j eventually lies in an A_{nk} with fixed k, or x_j is not covered by more than a finite number of $\{U_n\}$. By the diagonal process, we can get a subsequence such that this is true of all the atoms.

We can choose an increasing sequence $\{S_j\}$ of compact subsets of X such that

$$\lim_{j\to\infty} \mu(S_j) = 1$$

[4], p. 64. The closed sets in S_j form a compact metric space in the Hausdorff metric [2], pp. 166–172:

$$d^*(E, F) = \max\{\max_{x\epsilon E} \min_{y\epsilon F} d(x, y), \max_{y\epsilon F} \min_{x\epsilon E} d(x, y)\}.$$

Hence by two applications of the diagonal process, we can get a subsequence of $\{U_n\}$ such that in the metric d^*

$$B_{kj} = \lim_{n\to\infty} \overline{A_{nk} \cap S_j}$$

exists for all k, j. The explicit formula for B_{kj} is

$$B_{kj} = \bigcap_{r=1}^{\infty} \overline{(\bigcup_{n=r}^{\infty} A_{nk} \cap S_j)}.$$

The diameter of a set is a continuous function in the d^* metric. Hence

$$\operatorname{diam} (B_{kj}) = \lim_{n\to\infty} \operatorname{diam} (A_{nk} \cap S_j) \leqq \epsilon.$$

Also, it follows from the formula for the limit that

$$\bigcup_{k=1}^{r} B_{kj} = \lim_{n\to\infty} \overline{(\bigcup_{k=1}^{r} A_{nk}) \cap S_j},$$

and if ν_r is any countably additive measure defined on the open and closed sets,

$$\nu_r(\bigcup_{k=1}^r B_{kj}) \geqq \limsup_{n\to\infty} \sum_{k=1}^r \nu_r(\bar{A}_{nk}) - \nu_r(X - S_j).$$

Let Z be the set of atoms of X, and Z_k the set of atoms which are eventually contained in A_{nk} as $n \to \infty$. Define

$$Y_r = Z - \bigcup_{k=1}^r Z_k.$$

This set consists of the atoms which eventually do not lie in $A_{n1}, \cdots, A_{nr}$. Any finite subset of Y_r will not intersect $\bigcup_{k=1}^r A_{nk}$ if n is sufficiently large. Hence if we define

$$\nu_r(E) = \mu(E - Y_r),$$

then

$$\limsup_{n\to\infty} \sum_{k=1}^r \nu_r(\bar{A}_{nk}) \geqq \lim_{n\to\infty} \sum_{k=1}^r \mu(A_{nk}) = \sum_{k=1}^r p_k,$$

and

$$(5) \qquad \mu(\bigcup_{k=1}^r B_{kj} - Y_r) \geqq \sum_{k=1}^r p_k - \mu(X - S_j - Y_r).$$

Define

$$B_k = \bigcup_{j=1}^\infty B_{kj}.$$

$B_{k1}, B_{k2}, \cdots$ form an increasing sequence of sets. Hence

$$\text{diam}\,(B_k) = \lim_{j\to\infty} \text{diam}\,(B_{kj}) \leqq \epsilon.$$

Let $j \to \infty$ in (5). We get

$$\mu(\bigcup_{k=1}^r B_k - Y_r) \geqq \sum_{k=1}^r p_k.$$

Define $\{C_k\}$ inductively by

$$C_1 = B_1 - Y_1,$$
$$C_r = \bigcup_{k=1}^r B_k - Y_r - \bigcup_{k=1}^{r-1} C_k, \qquad r \geqq 2.$$

Then the C_k are disjoint ϵ-sets, and

$$(6) \qquad \sum_{k=1}^r \mu(C_k) \geqq \sum_{k=1}^r p_k.$$

Also, $Z_k \subset B_k$, so it is easy to show by induction that $Z \cap C_k = Z_k$.

It follows from (6) that

$$(7) \qquad \sum_{k=1}^\infty \mu(C_k) \geqq \sum_{k=1}^\infty p_k = m.$$

If equality holds here, define $\{C_k'\} = \{C_k\}$. Otherwise, we will define a sequence $\{C_k'\}$, with $C_k' \subset C_k$, such that

$$(6') \qquad \sum_{k=1}^r \mu(C_k') \geqq \sum_{k=1}^r p_k, \qquad r = 1, 2, \cdots,$$

$$(8) \qquad \sum_{k=1}^\infty \mu(C_k') = m.$$

We have $\mu(Z_k) \leqq p_k$. Hence if $\sum_{k=1}^\infty \mu(C_k) > m$, there is a first index k_1 such that

$$\sum_{k=1}^{k_1} \mu(C_k) + \sum_{k=k_1+1}^{\infty} \mu(Z_k) > m.$$

Then

$$\sum_{k=1}^{k_1-1} \mu(C_k) + \sum_{k=k_1}^{\infty} \mu(Z_k) \leqq m.$$

Define

$$C_k' = C_k, \quad k < k_1, \qquad C_k' = Z_k, \quad k > k_1.$$

Then

$$\mu(Z_{k_1}) \leqq m - \sum_{k \neq k_1} \mu(C_k') < \mu(C_{k_1}).$$

The set $C_{k_1} - Z_{k_1}$ is non-atomic. Hence it has a subset of any positive measure between zero and $\mu(C_{k_1} - Z_{k_1})$. It follows that we can choose C'_{k_1}, $Z_{k_1} \subset C'_{k_1} \subset C_{k_1}$, such that

$$\mu(C'_{k_1}) = m - \sum_{k \neq k_1} \mu(C_k').$$

Then (8) is satisfied. Now (6′) is equivalent to (6) for $r < k_1$. For $r \geqq k_1$, (6′) is implied by (8) and the inequality

$$\sum_{k=r}^{\infty} \mu(C_k') = \sum_{k=r}^{\infty} \mu(Z_k) \leqq \sum_{k=r}^{\infty} p_k, \qquad r > k_1.$$

Now Lemma 2 applies. If $U \rightleftharpoons \{C_k'\}$, by (4), (6′) and (8) we have

$$H(U) \leqq \sum (p_k/m) \log (m/p_k) \leqq L.$$

This completes the proof of Theorem 1.

3. Consequences of lower semicontinuity. It is the purpose of this section to derive some consequences of the preceding lower semicontinuity theorem. The first states that the infimum of (2) is actually achieved.

THEOREM 2. *For every $\epsilon > 0$ and $\delta \geqq 0$, there exists an $\epsilon; \delta$ partition U of X such that $H(U) = H_{\epsilon;\delta}(X)$.*

PROOF. In Theorem 1, let every ϵ_n be equal to ϵ, and every δ_n be equal to δ, where $\{U_n\}$ is a sequence of $\epsilon; \delta$ partitions of X such that

$$H_{\epsilon;\delta}(X) = \lim_{n\to\infty} H(U_n).$$

Theorem 1 produces an $\epsilon; \delta$ partition U of X with

$$H(U) = \lim_{n\to\infty} H(U_n),$$

as required. Theorem 2 is proved.

The following theorem is of interest in its own right.

THEOREM 3. *Let (X, d, μ) be a probabilistic metric space, let $0 < p \leqq 1$, and let $\epsilon \geqq 0$. Suppose that for every $p' < p$, there is a Borel set $B(p')$ of diameter at most ϵ such that $\mu(B(p')) \geqq p'$. Then there is a Borel set $B(p)$ of diameter at most ϵ such that $\mu(B(p)) \geqq p$.*

PROOF. First let $\epsilon = 0$. The hypothesis is that atoms exist of probability at least p', for every $p' < p$, and we are to produce an atom of probability at least p. But since the sum of the probabilities of the atoms of X is finite, there is an atom of maximum probability. The hypothesis implies that the probability of this atom is at least p.

Now let $\epsilon > 0$. Let p_n converge to p from below, and let $\{B(p_n)\}$ be Borel sets of diameter $\leqq \epsilon$ such that $\mu(B(p_n)) \geqq p_n$. In Theorem 1, let $\delta_n = 1 - p_n$, $\delta = 1 - p$. We are told that for every n, there is an $\epsilon;\delta_n$ partition U_n with $H(U_n) = 0$; namely, the partition $\{B(p_n)\}$ consisting of one set will do. By Theorem 1, there is an an $\epsilon; \delta$ partition U of X such that $H(U) = 0$.

But if $H(U) = 0$, then U consists of one Borel set of positive measure, which we call $B(p)$. Since U is an $\epsilon; \delta$ partition,

$$\text{diam}\,(B(p)) \leqq \epsilon, \qquad \mu(X - B(p)) \leqq \delta.$$

Thus $\mu(B(p)) \geqq p$. Theorem 3 is proved.

The next theorem states that $H_{\epsilon;\delta}(X)$ is jointly continuous in $\epsilon; \delta$ from above. We note that separate continuity in $\epsilon; \delta$ from above is enough to force joint continuity, since $H_{\epsilon;\delta}(X)$ is non-increasing in ϵ and δ.

THEOREM 4. *The function $H_{\epsilon;\delta}(X)$ is jointly continuous from above in ϵ and δ for $\epsilon > 0$, $\delta \geqq 0$.*

PROOF. Let ϵ_n converge to ϵ from above, and δ_n converge to δ from above. By Theorem 2, there exist $\epsilon_n; \delta_n$ partitions U_n such that

$$H(U_n) = H_{\epsilon_n;\delta_n}(X).$$

By Theorem 1, there is an $\epsilon; \delta$ partition U of X such that

$$H(U) \leqq \liminf_{n\to\infty} H(U_n).$$

Thus,

$$H_{\epsilon;\delta}(X) \leqq \lim_{\epsilon'\to\epsilon+,\delta'\to\delta+} H_{\epsilon';\delta'}(X).$$

The inequality

$$H_{\epsilon;\delta}(X) \geqq \lim_{\epsilon'\to\epsilon+,\delta'\to\delta+} H_{\epsilon';\delta'}(X)$$

follows from the fact that the function $H_{\epsilon;\delta}(X)$ is monotone non-increasing in ϵ and δ. Thus

$$H_{\epsilon;\delta}(X) = \lim_{\epsilon'\to\epsilon+,\delta'\to\delta+} H_{\epsilon';\delta'}(X),$$

which proves Theorem 4.

Continuity of $H_{\epsilon;\delta}(X)$ from below in δ does not necessarily hold. For example, let X be the two-point probabilistic metric space, where each probability is $\frac{1}{2}$ and the two points are at distance 1. Then $H_{\frac{1}{2};\frac{1}{2}}(X) = 0$, whereas $\lim_{\delta\to\frac{1}{2}-} H_{\frac{1}{2};\delta}(X) = \log 2$. However, the following theorem holds.

THEOREM 5. *Let X either be non-atomic or have infinite ϵ-entropy, for given $\epsilon > 0$. Let $0 \leqq \delta \leqq 1$. Then $H_{\epsilon;\delta}(X)$ is continuous in δ from below.*

PROOF. First let X be non-atomic. Let $\eta > 0$ be given. We shall find a $\delta' < \delta$ such that for every ρ, $\delta' \leqq \rho \leqq \delta$, there is an $\epsilon; \rho$ partition U of X with $H(U) \leqq H_{\epsilon;\delta}(X) + \eta$. Thus, $\lim_{\rho\to\delta-} H_{\epsilon;\rho}(X) \leqq H_{\epsilon;\delta}(X)$, and $H_{\epsilon;\delta}(X)$ will indeed be continuous in δ from below.

Let V be an $\epsilon; \delta$ partition of X such that $H(V) \leqq H_{\epsilon;\delta}(X) + \eta/2$. If $\mu(V) = 1$, V is already the type of $\epsilon; \rho$ partition desired, for any $\rho \geqq 0$. Otherwise, let A be the set (of positive probability $1 - \mu(V) \leqq \delta$) not covered by any set in V. Since A contains no atoms, A contains a set B_λ of diameter $\leqq \epsilon$ and positive probability $\lambda \leqq \lambda_0$, for any $\lambda_0 > 0$ (merely take a set of positive probability from any sufficiently fine partition of A). Choose $\lambda_0 > 0$ so small that if $U = V \cup \{B_\lambda\}$, $|H(U) - H(V)| < \eta/2$. Then $H(U) \leqq H_{\epsilon;\delta}(X) + \eta$. And of course U is an $\epsilon; \delta - \lambda$ partition of X. Let $\delta' = \delta - \lambda$ to conclude that U is an $\epsilon; \rho$ partition of X for every $\delta' \leqq \rho \leqq \delta$. This proves continuity from below in case X has no atoms.

Now let $H_\epsilon(X)$ be infinite. That is, let there be no ϵ-partition of X of finite entropy. As above, choose an $\epsilon; \delta$ partition V and assume that its uncovered set A has positive probability $\leqq\delta$. We need to show that for every $\lambda_0 > 0$, A contains a set of probability greater than zero but $\leqq\lambda_0$ and of diameter $\leqq\epsilon$. If for some $\lambda_0 > 0$ every set of diameter $\leqq\epsilon$ of positive probability in A had probability $>\lambda_0$, then A could be covered by finitely many sets of diameter $\leqq\epsilon$, say B_1, $B_2, \cdots, B_k$. Then $U = V \cup \{B_1, \cdots, B_k\}$ would be an ϵ-partition of X of finite entropy. Theorem 5 is proved.

We observe that not even the absence of atoms makes $H_{\epsilon;\delta}(X)$ continuous from below in ϵ. For example, let X be the circle of radius 1 in the Euclidean plane, let d be Euclidean distance, and let μ be angular measure divided by 2π. Then $H_2(X) = 0$, whereas $\lim_{\epsilon\to 2-} H_\epsilon(X) = \log 2$. The latter is true because when $\epsilon < 2$, no set in an ϵ-partition of X can contain a pair of antipodal points. However, it is shown in [3] that $H_\epsilon(X)$ is indeed continuous from below when X is the probabilistic metric space consisting of $L_2[0, 1]$ under L_2-norm, where the measure is that induced on $L_2[0, 1]$ by a mean-continuous Gaussian process. Since such X are non-atomic, we conclude from Theorem 4, together with Theorem 5 and the remark preceding it, that $H_{\epsilon;\delta}(X)$ is jointly continuous in ϵ and δ for X equal to $L_2[0, 1]$ under the measure induced by a mean-continuous Gaussian process.

Let us now define $H_{\epsilon;\delta}(X)$ (and $H_\epsilon(X)$) for $\epsilon = 0$ as follows: $H_{0;\delta}$ is infinite if there is no partition of at least $1 - \delta$ of X by sets of diameter 0 (atoms); if there is such a partition, the definition is the same as before. With this definition, $H_{0;\delta}(X)$ is continuous in δ from above, as is easy to see. And $H_{\epsilon;\delta}(X)$ is continuous from above in ϵ as $\epsilon \to 0$, δ fixed, as Theorem 1 shows (a simpler proof suffices, not using the Hausdorff metric d^*). The following result is also true:

THEOREM 6. *If the set Z of atoms of X has measure $1 - \delta_0$, then $H_{0;\delta_0}(X) = H(Z)$ and $H_{0;\delta}(X)$ is finite for $\delta > \delta_0$, infinite for $\delta < \delta_0$.*

PROOF. Any $0; \delta_0$ partition of X has as its sets all the atoms, and perhaps other sets of probability zero; hence its entropy is $H(Z)$. For $\delta < \delta_0$, there are no admissible 0; δ-partitions, while for $\delta > \delta_0$, there are finite 0; δ partitions. Thus all the statements of the theorem follow.

The following theorem is of interest, but is to be expected: one should expect to need arbitrarily many bits to prescribe a nonatomic probabilistic metric space with arbitrary precision.

THEOREM 7. *If X is non-atomic, $H_{0;\delta}(X)$ is infinite for every $\delta < 1$. Thus, $H_{\epsilon;\delta}(X) \to \infty$ as $\epsilon \to 0$, for $\delta < 1$. If X is not purely atomic, $H_\epsilon(X) \to \infty$ as $\epsilon \to 0$.*

PROOF. Combine Theorems 4 and 6.

4. Finite-dimensional euclidean spaces. As preparation for the final section, this section considers the ϵ-entropy of distributions in Euclidean n-space (with ordinary sum-of-squares metric). This certainly yields a probabilistic metric space. We assume that $E(\|x\|^2) < \infty$, where E denotes expectation, $\|x\|$ denotes Euclidean norm. We can prove

THEOREM 8. *Let X be Euclidean n-space with a probability distribution such that $\|x\|^2$ has finite expectation. Then $H_\epsilon(X)$ is finite, for every $\epsilon > 0$. In fact, $H_\epsilon(X) \leqq n \log (1/\epsilon) + C$ for small ϵ, where the constant C depends only on n and $E\{\|x\|^2\}$.*

We first need three lemmas.

LEMMA 3. *Let X be a product space,*

$$X = \textstyle\prod_{j=1}^{J} X_j ,$$

where J is a positive integer or ∞. Let there be some probability distribution on X, inducing marginal distributions on each X_j. Let U_j be a denumerable partition of X_j for each j, and consider the product partition *U of X given as*

$$U = \{\textstyle\prod_{j=1}^{J} A_j\} = \textstyle\prod_{j=1}^{J} U_j ,$$

where each $A_j \in U_j$. Assume all the sets are measurable. Then if

$$\textstyle\sum_{j=1}^{J} H(U_j) < \infty,$$

all but a denumerable number of the sets of U lie in a subset of X of probability zero, and

$$H(U) \leqq \textstyle\sum_{j=1}^{J} H(U_j). \tag{9}$$

More specifically, if $U_j = \{U_{jk}\}$, with the sets arranged in order of non-increasing probability, all sets of U which involve infinitely many U_{jk} with $k > 1$ lie in a set of probability zero.

PROOF. Suppose first that J is finite. Then U is denumerable, and the only thing to prove is (9). For $x_j \varepsilon X_j$, let z_j be the index of the set of U_j to which x_j belongs. Then $z_1, \cdots, z_j$ are discrete random variables, and $(z_1, \cdots, z_J)$ is a vector random variable designating the set of U containing $x \varepsilon X$. Inequality (9) is equivalent to $H(z_1, \cdots, z_J) \leqq \sum_{j=1}^{J} H(z_j)$, which is one of the basic inequalities of Information Theory. ([1], p. 16; the proof there is easily extended to variables which take an infinite number of values.)

Now let $J = \infty$. First we prove the last statement of the lemma. For each j, U_{j1} is one of the most likely sets of U_j. Hence $\mu(U_{j1}) \geqq e^{-H(U_j)}$, and $\mu(X_j - U_{j1}) \leqq 1 - e^{-H(U_j)} \leqq H(U_j)$. Summing,

$$\sum_{j=1}^{\infty} \mu(X_j - U_{j1}) \leqq \sum_{j=1}^{\infty} H(U_j) < \infty .$$

The Borel-Cantelli lemma then implies that the set of points of X whose projections into the X_j lie outside an infinite number of the U_{j1}'s has probability zero.

It now follows that

$$\begin{aligned} H(U) &= \lim_{N\to\infty} H(U_1 \times U_2 \times \cdots \times U_N) \\ &\leqq \lim_{N\to\infty} \sum_{j=1}^{N} H(U_j) = \sum_{j=1}^{\infty} H(U_j), \end{aligned}$$

by the result for finite J. This completes the proof of the lemma.

REMARK. This lemma can be applied to separable Hilbert space, by embedding it in the (non-metric) space which is the product of the one-dimensional coordinate spaces, forming the product partition, and restricting the product sets to Hilbert space.

LEMMA 4. *Let* $\{p_1, p_2, \cdots\}$ *and* $\{q_1, q_2, \cdots\}$ *be sequences of non-negative numbers with*

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} q_k \leqq 1.$$

Then

$$\sum_{k=1}^{\infty} p_k \log (1/p_k) \leqq \sum_{k=1}^{\infty} p_k \log (1/q_k)$$

(*any term with* $p_k = 0$ *is interpreted as zero;* $p_k \log (1/q_k) = \infty$ *if* $q_k = 0 \neq p_k$).

PROOF. If $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$ are two sets of non-negative numbers with $\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} b_k = 1$, we have

$$\sum_{k=1}^{n} a_k \log (1/a_k) \leqq \sum_{k=1}^{n} a_k \log (1/b_k)$$

[1], p. 13. For any positive number c, replacing a_k, b_k by ca_k, cb_k changes both expressions in such a way that the inequality is preserved. Hence $\sum a_k$ and $\sum b_k$ can have any common positive value.

Apply this inequality to

$$\begin{aligned} a_k &= p_k , & b_k &= q_k , & k &\leqq n - 1, \\ a_n &= 1 + \sum_{k=n}^{\infty} p_k , & b_n &= 1 + \sum_{k=n}^{\infty} q_k . \end{aligned}$$

We get

$$\sum_{k=1}^{n-1} p_k \log (1/p_k) + a_n \log (1/a_n) \leqq \sum_{k=1}^{n-1} p_k \log (1/q_k) + a_n \log (1/b_n).$$

As $n \to \infty$, a_n, $b_n \to 1$. Hence the last terms on each side of this inequality approach zero, and the desired inequality is obtained.

LEMMA 5. *Let a real vector-valued random variable* ξ *have finite second moment* σ^2, i.e., $E(\|\xi\|^2) = \sigma^2$. *If*

$$p_k = \mu\{k\epsilon \leqq \|\xi\| < (k + 1)\epsilon\}, \qquad k = 0, 1, 2, \cdots ,$$

then

(10) $$H = \sum_{k=0}^{\infty} p_k \log (1/p_k) \leqq \tfrac{2}{3}e \log (e\sigma^2/\epsilon^2), \qquad \epsilon \leqq \sigma,$$
$$\leqq \tfrac{2}{3}e(\sigma^2/\epsilon^2) \log (e\epsilon^2/\sigma^2), \qquad \sigma \leqq \epsilon.$$

Also, for $\epsilon \leqq \sigma$,

(11) $$H \leqq \log (\sigma/\epsilon) + C,$$

where C is a universal constant.

PROOF. Let α be any positive number, and $A = \sum_{j=0}^{\infty} e^{-\alpha j^2}$. By hypothesis, $\epsilon^2 \sum k^2 p_k \leqq \sigma^2$. Hence if we apply Lemma 4 to the sequence $\{p_k\}$ and $\{e^{-\alpha k^2}/A, k \geqq 0\}$, the result is

$$\sum_{k=0}^{\infty} p_k \log (1/p_k) \leqq \sum_{k=0}^{\infty} p_k(\alpha k^2 + \log A) \leqq \alpha\sigma^2/\epsilon^2 + \log A = F(\alpha, \epsilon/\sigma).$$

The bounds to be established are bounds on the function

$$G(\epsilon/\sigma) = F(\epsilon^2/\sigma^2, \epsilon/\sigma), \qquad \epsilon \leqq \sigma,$$
$$= F(1 + 2 \log \epsilon/\sigma, \epsilon/\sigma), \qquad \epsilon > \sigma.$$

For $t \leqq 1$, we have

$$\sum_{j=0}^{\infty} e^{-j^2 t^2} < 1 + \int_0^{\infty} e^{-t^2 u^2}\, du = 1 + \pi^{\frac{1}{2}}/2t,$$

hence

(12) $$G(t) \leqq 1 + \log (1 + \pi^{\frac{1}{2}}/2t) \leqq \log (1/t) + 1 + \log (1 + \pi^{\frac{1}{2}}/2),$$

verifying (11). For $t \geqq 1$,

$$\sum_{j=0}^{\infty} e^{-j^2 \log(et^2)} = \sum_{j=0}^{\infty} (et^2)^{-j^2} \leqq et^2/(et^2 - 1) \leqq 1 + 1/(e - 1)t^2;$$

hence

(13) $$G(t) \leqq (1 + 2 \log t)/t^2 + 1/(e - 1)t^2 = (1/t^2)(e/(e - 1) + 2 \log t).$$

From (12) and (13), it follows that (10) is valid, since

$$\max \{e/(e - 1), 1 + \log (1 + \pi^{\frac{1}{2}}/2)\} < \tfrac{2}{3}e.$$

This completes the proof of Lemma 5.

PROOF OF THEOREM 8. First let $n = 1$. One possible ϵ-partition is formed by decomposing the line of values of x into intervals of length ϵ, with end-points at $0, \pm\epsilon, \pm 2\epsilon, \cdots$. This partition U is a refinement of the partition of Lemma 5, decomposing each set of probability p_k into two sets of probability q_k, r_k. By Lemma 4,

$$q_k \log (1/q_k) + r_k \log (1/r_k) \leqq q_k \log (2/p_k) + r_k \log (2/p_k)$$
$$= p_k \log (1/p_k) + p_k \log 2.$$

Hence, applying Lemma 5,

$$H(U) = \sum_{k=0}^{\infty} [q_k \log (1/q_k) + r_k \log (1/r_k)] \leqq \sum_{k=0}^{\infty} [p_k \log (1/p_k) + p_k \log 2]$$
$$\leqq \log (\sigma/\epsilon) + C + \log 2$$

for $\epsilon \leqq \sigma$. Theorem 8 is proved for $n = 1$.

If $n > 1$, let U_j be the partition of this type of the jth coordinate line, using intervals of length $\epsilon/n^{\frac{1}{2}}$. If $\epsilon \leqq n \min [E(x_j^2)]^{\frac{1}{2}}$, we have, from above,

$$H(U_j) \leqq \log \{(1/\epsilon)[nE(x_j^2)]^{\frac{1}{2}}\} + C + \log 2.$$

The product partition U formed from $\{U_j\}$ is an ϵ-partition of n-space. By Lemma 3

$$\begin{aligned} H(U) &\leqq \textstyle\sum_{j=1}^{\infty} (\log \{(1/\epsilon)[nE(x_j^2)]^{\frac{1}{2}}\} + C + \log 2) \\ &\leqq n \log (1/\epsilon) + n\{\log [nE\{\|x\|^2\}]^{\frac{1}{2}} + C + \log 2\}. \end{aligned}$$

This completes the proof of Theorem 8.

REMARK. In [5], sharper results on the ϵ-entropy of finite-dimensional distributions in Euclidean spaces are proved, for example in the case in which the distribution is absolutely continuous with respect to Lebesgue measure with a continuous density function. In particular, the relation between ϵ-entropy and the "differential entropy" ($\int p \log (1/p)\, dx$ for p a density function) is pointed out.

5. Mean-continuous stochastic processes. The first theorem of this final section can be regarded as a generalization of Theorem 8 to countable-dimensional Euclidean spaces, in particular, to $L_2[0, 1]$. Let $Y(t, \omega)$, $\omega \in \Omega$, be a separable mean-continuous stochastic process on the closed unit interval (see [3], Chapter X for definitions). Let $R(s, t) = E(Y(s, \omega)Y(t, \omega))$, a continuous positive-definite function. Let $\sigma_1^2 \geqq \sigma_2^2 \geqq \cdots \geqq 0$ be the eigenvalues of the integral equation $\int R(s, t)\phi(t)dt = \sigma^2\phi(s)$, with associated normalized continuous eigenfunctions $\phi_n(t)$. (It is known that $R(s, t) = \sum \sigma_n^2\phi_n(s)\phi_n(t)$ uniformly on the unit square; also, $\sum_{n=1}^{\infty} \sigma_n^2$ is finite and equal to $E(\int_0^1 Y^2(t, \omega)\, dt)$.) Then there exist uncorrelated random variables $\xi_n(\omega)$, $E(\xi_n^2) = 1$, such that with probability 1 on Ω

$$\textstyle\sum_{n=1}^{N} \sigma_n\xi_n(\omega)\phi_n(t) \rightarrow Y(t, \omega) \quad \text{in} \quad L_2[0, 1]$$

as $N \rightarrow \infty$.

Using the techniques of [7], what we really have is a probability distribution μ on an abstract separable Hilbert space with orthonormal basis $\{\phi_n\}$ (the form of $\{\phi_n\}$ has nothing to do with the L_2-structure of the process). The measure μ is uniquely determined from the finite-dimensional distributions of $(\xi_1, \xi_2, \cdots, \xi_n)$, all n. The distribution in this n-space is that of $\sum_{j=1}^{n} \sigma_j\xi_j\phi_j$. All other facts about the process can be ignored in what follows.

For convenience, we call the ϵ-entropy of the probabilistic metric space X, consisting of $L_2[0, 1]$ under L_2-norm with measure μ, the *ϵ-entropy of the process.* It might be reasoned by analogy with Theorem 8 that, since $\sum \sigma_n^2 < \infty$, this ϵ-entropy should be finite for every positive ϵ. Such however is not necessarily the case. (Don't try to find an example of a mean-continuous process which is Gaussian with infinite ϵ-entropy for some $\epsilon > 0$, though. It is shown in [6] that the ϵ-entropy of a Gaussian process is always finite.) We need three preliminary lemmas before getting to Theorem 9.

LEMMA 6. *Let a maximal set of points on the unit sphere in n-space* $(n \geqq 1)$ *with mutual distances* $\geqq \delta$ $(\delta \leqq 1)$ *contain N points. Then*

$$\delta^{-n+1} \leqq N \leqq 4^n \delta^{-n}.$$

PROOF. If $n = 1$, the inequality is trivial. Henceforth assume $n \geqq 2$. The area of a cap of radius r on the sphere $(r \leqq 1)$ can be given an integral formula by projecting onto an equatorial hyperplane and using spherical coordinates in $(n - 1)$-space. If π_n is the area of the unit sphere in n-space, the cap has area

$$A(r) = \pi_{n-1} \int_0^{r'} t^{n-2}\, dt/(1 - t^2)^{\frac{1}{2}},$$

where $r' = r(1 - \frac{1}{4}r^2)^{\frac{1}{2}}$. Replacing $(1 - t^2)^{\frac{1}{2}}$ by its upper and lower bounds, then integrating,

$$[\pi_{n-1}/(n - 1)](r')^{n-1} \leqq A(r) \leqq [\pi_{n-1}/(n - 1)] \cdot (r')^{n-1}/(1 - \tfrac{1}{2}r^2).$$

Our set has the property that the closed δ-neighborhoods of the points cover the sphere, while the open $\delta/2$-neighborhoods are disjoint. Hence $NA(\delta/2) \leqq \pi_n \leqq NA(\delta)$. Using the above bounds on $A(r)$,

$$g_n[\delta(1 - \delta^2/4)^{\frac{1}{2}}]^{-n+1}(1 - \tfrac{1}{2}\delta^2) \leqq N \leqq g_n[\tfrac{1}{2}\delta(1 - \delta^2/16)^{\frac{1}{2}}]^{-n+1},$$

where $g_n = (n - 1)\pi_n/\pi_{n-1} = 2\pi^{\frac{1}{2}}\Gamma((n + 1)/2)/\Gamma(n/2)$. Since $\delta \leqq 1$, we have $\frac{1}{2}g_n\, \delta^{-n+1} \leqq N \leqq g_n(\delta(15)^{\frac{1}{2}}/8)^{-n+1}$.

From the relation $g_{n+2} = (1 + 1/n)g_n$, together with $g_1 = 2, g_2 = \pi$, it follows easily that $2 \leqq g_n \leqq 2((15)^{\frac{1}{2}}/2)^{n-1}$. Hence

$$\delta^{-n+1} \leqq N \leqq 2(\delta/4)^{-n+1} < (4/\delta)^n.$$

LEMMA 7. *The ϵ-entropy of any distribution over an $(n - 1)$-sphere X of radius ρ in n-dimensional Euclidean space is at most* $n \log^+ (8\rho/\epsilon)$.

PROOF. The result is clearly true for $\epsilon \geqq 2\rho$, for then $H_\epsilon(X) = 0$. Suppose $\epsilon < 2\rho$. Take a maximal set of points in X at distances at least $\epsilon/2$. If the set contains N points, then, by Lemma 6, $N \leqq (8\rho/\epsilon)^n$. The closed $\epsilon/2$-neighborhoods of these points form an ϵ-covering of X, from which an ϵ-partition can be extracted. Since there is an ϵ-partition containing at most N sets, $H_\epsilon(X) \leqq \log N \leqq n \log (8\rho/\epsilon)$. This proves Lemma 7.

LEMMA 8. *Suppose* $\sum_{k=1}^\infty 2^k \rho_k^2 = S < \infty$. *Then there is a sequence* $\{\epsilon_k\}$, with $\sum_{k=1}^\infty \epsilon_k^2 = \epsilon^2$, *such that*

$$H = \sum_{k=1}^\infty 2^k \log^+ \rho_k^2/\epsilon_k^2 < 5S/\epsilon^2.$$

PROOF. We can choose $\{\epsilon_k\}$ so that eventually $\epsilon_k \geqq \rho_k$, and the series for H terminates. If any $\epsilon_k > \rho_k$, then decreasing this ϵ_k while increasing an $\epsilon_{k'} < \rho_{k'}$ decreases H. Hence, we may suppose that $\epsilon_k < \rho_k$ on some finite set of indices P, and $\epsilon_k = \rho_k$ on the complement P'.

We have $H = \sum_P 2^k \log (\rho_k^2/\epsilon_k^2)$, where these ϵ_k's must satisfy the condition $\sum_P \epsilon_k^2 = \epsilon^2 - \sum_{P'} \epsilon_k^2$. By varying the ϵ_k subject to this condition, it is seen that H is maximum when $\epsilon_k^2 = 2^k A$, where A is some constant. It follows that the best ϵ_k's have the form $\epsilon_k^2 = \min(2^k A, \rho_k^2)$. Put $\rho_k^2 = 2^k A g_k$. Then

$$g_k \geqq 1 \quad \text{on} \quad P, \qquad g_k \leqq 1 \quad \text{on} \quad P',$$

and

$$A \sum_{k=1}^{\infty} 4^k g_k = S, \tag{14}$$

$$A\left(\sum_P 2^k + \sum_{P'} 2^k g_k\right) = \epsilon^2, \qquad H = \sum_P 2^k \log g_k. \tag{15}$$

When $g_k = 1$, k can be put in either P or P', whichever is convenient.

For variations of $\{g_k\}$ in P subject to (14), we find that H is maximum when $g_k = \max(2^{-k}B, 1)$, $k \varepsilon P$, where B is a constant. Putting k in P' whenever $g_k \leqq 1$ and eliminating A from (14) and (15), we have

$$\epsilon^2\left(\sum_P 2^k B + \sum_{P'} 4^k g_k\right) = S\left(\sum_P 2^k + \sum_{P'} 2^k g_k\right), \tag{16}$$

while $H = \sum_P 2^k \log(2^{-k}B)$. Vary $g_{k_1}(k_1 \varepsilon P')$ and B, subject to (16). We have

$$\epsilon^2\left(\sum_P 2^k\right)\delta B = (S - 2^{k_1}\epsilon^2)\cdot 2^{k_1}\delta g_{k_1}.$$

A variation of g_k which increases B increases H. Hence we may assume that

$$g_k = 0 \quad \text{for} \quad 2^k\epsilon^2 \geqq S,$$

$$g_k = 1 \quad \text{for} \quad 2^k\epsilon^2 < S,$$

for $k \varepsilon P'$.

Now, all k's for which $g_k \neq 0$ can be put into P. Thus, we need only consider the case

$$g_k = \max(2^{-k}B, 1), \qquad k \varepsilon P,$$

$$g_k = 0, \qquad k \not\varepsilon P.$$

We have

$$\epsilon^2 \sum_P \max(2^k B, 4^k) = S \sum_P 2^k; \tag{17}$$

also, $H = \sum_P 2^k \log^+(2^{-k}B)$.

Let K be the last index for which there is a non-zero contribution to H. Then

$$\begin{aligned} H \leqq \sum_{k=1}^{K} 2^k \log(2^{-k}B) &\leqq \sum_{m=0}^{\infty} 2^{K-m}[\log(B/2^K) + m \log 2] \\ &= 2^{K+1}[\log(B/2^K) + \log 2]. \end{aligned} \tag{18}$$

If K is the last index in P, then from (17) we have $2^K B\epsilon^2 \leqq 2^{K+1}S$ and $2^K B \geqq 4^K$. Thus $2^K \leqq B \leqq 2S/\epsilon^2$ which makes

$$\begin{aligned} H &\leqq (4S/\epsilon^2)\cdot(2^K/B)\log(B/2^K) + (4S/\epsilon^2)\log 2 \\ &\leqq (4S/\epsilon^2)(e^{-1} + \log 2), \end{aligned} \tag{19a}$$

since the function $x \log(1/x)$ has maximum value $1/e$.

The case remains in which the last indices of P have $B < 2^K$. Then the number K in (18) can be increased, if necessary, so that $2^K \leqq B \leqq 2^{K+1}$, and, if L is the last index in P, $K < L$. Thus $H \leqq 2^{L+1} \log 2$. From (17),

$$4^L \epsilon^2 \leqq 2^{L+1} S, \qquad 2^L \leqq 2S/\epsilon^2.$$

Hence

$$H \leqq (4S/\epsilon^2) \log 2. \tag{19b}$$

Since the maximum of H satisfies either (19a) or (19b), and since $1/e + \log 2 < \frac{5}{4}$, we certainly have $H < 5S/\epsilon^2$.

Lemma 8 is proved, and we can return to the statement and proof of Theorem 9.

THEOREM 9. *If $\sum n\sigma_n^2 < \infty$, any mean-continuous stochastic process on* [0, 1] *with eigenvalues $\{\sigma_n^2\}$ has finite ϵ-entropy H_ϵ. Specifically,*

$$H_\epsilon = o(1/\epsilon^2) \quad as \quad \epsilon \to 0,$$

where the function o depends only on the eigenvalues σ_n^2 of the process.

PROOF. We actually prove that if $\epsilon^2 \leqq \sum n\sigma_n^2$, then

$$H_\epsilon < 12^3 \sum n\sigma_n^2/\epsilon^2. \tag{20}$$

We then use Theorem 8 as follows. Let N be so large that $\sum_{n>N} n\sigma_n^2 < \eta$, η a given positive number. Then the process which is related to the original process by ignoring the first N eigenfunctions has eigenvalues $(\sigma'_n)^2$ with $\sum_{n=1}^{\infty} n(\sigma'_n)^2 < \eta$. Hence, for $\epsilon^2/2 \leqq \sum n(\sigma'_n)^2$, the $\epsilon/2^{\frac{1}{2}}$-entropy of the related process is less than $3456\eta/\epsilon^2$. The N-dimensional subspace has $\epsilon/2^{\frac{1}{2}}$-entropy bounded in Theorem 8. The ϵ-entropy of the entire process is bounded, as in Theorem 8, by the sum of the entropies of product $\epsilon/2^{\frac{1}{2}}$-partitions obtained from the two factor spaces in question, as in Lemma 3. Hence, for sufficiently small ϵ depending only on the σ_n^2, we can guarantee that

$$H_\epsilon \leqq 3456\eta/\epsilon^2 + N \log (2^{\frac{1}{2}}/\epsilon) + C,$$

where C depends only on the σ_n^2. So for small enough ϵ depending only on the σ_n^2, we can guarantee that $H_\epsilon \leqq 3457\ \eta/\epsilon^2$. We conclude that H_ϵ is $o(1/\epsilon^2)$ as $\epsilon \to 0$, where o depends only on the eigenvalues of the process. Hence it is indeed enough to prove that for $\epsilon^2 \leqq \sum n\sigma_n^2$, (20) is valid. This we now proceed to do.

The idea is to group the coordinates into successive longer and longer finite-dimensional subspaces. Each such subspace is partitioned into "spherical shell regions" of constant radial separation in each space. However, the radial separation depends upon the space. The entropy of the product of these shell partitions is shown to be finite. In the nth finite-dimensional space, we subdivide each shell region by a partition whose entropy is estimated by projecting on to the inner sphere. The parameters involved in these partitions are so chosen that the partition of all of $L_2[0, 1]$ obtained from taking the union of the products of the partitions of the shell regions is an ϵ-partition of finite entropy. The previous lemmas allow the choices of parameters to be made so that the required facts can be proved.

Recall that the σ_n's are non-increasing. For $k = 1, 2, \cdots$, group the coordinates $\xi_{2^{k-1}}, \cdots, \xi_{2^k-1}$ into a vector y_k of 2^{k-1} components. Then y_k has second moment a_k^2 given by

$$E(\|y_k\|^2) = a_k^2 = \sum_{n=2^{k-1}}^{2^k-1} \sigma_n^2;$$

and

$$\sum n\sigma_n^2 \leqq S = \sum_{k=1}^{\infty} 2^k a_k^2 \leqq 2 \sum n\sigma_n^2. \tag{21}$$

Let $\{\delta_k\}$ be a sequence of positive numbers with $\sum \delta_k^2 \leqq \frac{1}{9}\epsilon^2$. We will first consider the partition U of the space formed by taking the product of a sequence of partitions of the y_k; the 2^{k-1}-dimensional space of y_k is subdivided into the *spherical shell regions*

$$l\delta_k \leqq \|y_k\| < (l+1)\delta_k, \qquad l = 0, 1, 2, \cdots.$$

The sequence $\{\delta_k\}$ will be chosen so that if H_k is the entropy of the kth component partition, $\sum H_k < \infty$. It will follow by Lemma 3 that all but a denumerable number of sets of U lie in a region of probability zero, and $H(U) \leqq \sum_{k=1}^{\infty} H_k$.

By Lemma 5,

$$H_k \leqq \tfrac{2}{3}e \log (ea_k^2/\delta_k^2), \qquad a_k \geqq \delta_k,$$

$$H_k \leqq \tfrac{2}{3}e(a_k^2/\delta_k^2) \log (e\delta_k^2/a_k^2), \qquad a_k \leqq \delta_k.$$

Put $\delta_k^2 = (\epsilon^2/9S)\cdot 2^k a_k^2$. Then

$$H(U) \leqq \sum_{2^k\epsilon^2 \leqq 9S} \tfrac{2}{3}e \log (9eS/2^k\epsilon^2) + \sum_{2^k\epsilon^2 > 9S} \tfrac{2}{3}e(9S/2^k\epsilon^2) \log (2^k e\epsilon^2/9S).$$

If $2\epsilon^2 \geqq 9S$,

$$\begin{aligned} H(U) &\leqq (6eS/\epsilon^2)\sum_{k=1}^{\infty} 2^{-k}[\log (e\epsilon^2/9S) + k\log 2) \\ &= (6eS/\epsilon^2) \log (4e\epsilon^2/9S). \end{aligned} \tag{22}$$

If $2\epsilon^2 \leqq 9S$, the case of interest, let the integer $K \geqq 1$ be such that $2^{k+\theta}\epsilon^2 = 9S$, with $0 \leqq 0 < 1$. Then

$$\begin{aligned} H(U) \leqq{}& \sum_{k=1}^{K} \tfrac{2}{3}e[\log (9eS/\epsilon^2) - k \log 2] \\ &+ \sum_{k=K+1}^{\infty} \tfrac{2}{3}e[(9S/2^k\epsilon^2) \log (e\epsilon^2/9S) + k \log 2] \\ \leqq{}& \tfrac{2}{3}eK \log (9eS/\epsilon^2) - \tfrac{1}{3}eK(K+1) \log 2 + (6eS/2^k\epsilon^2) \\ &\cdot[\log (e\epsilon^2/9S) + (K+2) \log 2] \\ \leqq{}& (2e/3 \log 2) \log (9S/\epsilon^2) \log (9eS/\epsilon^2) + \tfrac{4}{3}e(1 + \log 2), \end{aligned}$$

which certainly implies

$$H(U) \leqq 4e[\log (9eS/\epsilon^2)]^2. \tag{23}$$

Let the sets of positive probability in U be

$$A_j = \{l_{jk}\delta_k \leqq \|y_k\| < (l_{jk}+1)\delta_k, k = 1, 2, \cdots\}, \qquad j = 1, 2, \cdots.$$

If V_j is any ϵ-partition of A_j, then $V = \bigcup_j V_j$ is an ϵ-partition of the space, and, by the well-known formula for conditional entropy [9],

$$H(V) = H(U) + \sum \mu(A_j)H(V_j).$$

To construct V_j, we can take any $\epsilon/3$-partition $W_j = \{C_{jm}\}$ of the set $S_j = \{\|y_k\| = l_{jk}\delta_k,\ k = 1, 2, \cdots\}$, and define $C'_{jm} \,\varepsilon\, V_j$ to be the set of all points of A_j which project into points of C_{jm} when each y_k has its length decreased so that $\|y_k\| = l_{jk}\delta_k$. For, if $\{y_k\}$ projects into $\{y'_k\}$,

$$\sum \|y_k - y'_k\|^2 \leqq \sum \delta_k^2 = \epsilon^2/9.$$

Hence, if $\{z_k\}$ lies in the same C'_{jm} as $\{y_k\}$ and projects into $\{z'_k\}$,

$$\left(\sum \|z_k - y_k\|^2\right)^{\frac{1}{2}} \leqq \left(\sum \|z_k - z'_k\|^2\right)^{\frac{1}{2}} + \left(\sum \|z'_k - y'_k\|^2\right)^{\frac{1}{2}} + \left(\sum \|y'_k - y_k\|^2\right)^{\frac{1}{2}} \leqq \epsilon.$$

Thus, each V_j is an ϵ-partition of A_j.

If V_j is formed in this way, $H(V_j)$ can be made arbitrarily close to the $\epsilon/3$-entropy of the distribution on S_j induced by the projection. Thus

$$(24) \qquad H_\epsilon \leqq \inf_{\text{fixed } U} H(V) = H(U) + \sum \mu(A_j)H_{\epsilon/3}^{(j)},$$

where $H_{\epsilon/3}^{(j)}$ is the $\epsilon/3$-entropy of this distribution on S_j.

For fixed j, let $\{\eta_{jk}\}$ be some positive sequence of numbers with $\sum_{k=1}^{\infty} \eta_{jk}^2 = \frac{1}{9}\epsilon^2$, to be chosen later. Then if h_{jk} is the η_{jk}-entropy of the marginal distribution of y'_k on the sphere $\|y'_k\| = \rho_{jk} = l_{jk}\delta_k$, we have $H_{\epsilon/3}^{(j)} \leqq \sum_{k=1}^{\infty} h_{jk}$.

By Lemma 7, we have

$$H_{\epsilon/3}^{(j)} \leqq \sum_{k=1}^{\infty} 2^{k-1} \log^+ (8\rho_{jk}/\eta_{jk}) = \tfrac{1}{4}\sum_{k=1}^{\infty} 2^k \log^+ (64\rho_{jk}^2/\eta_{jk}^2).$$

It follows from Lemma 8 (using $8\rho_{jk}$ for ρ_k, η_{jk} for ϵ_k, and $\epsilon^2/9$ instead of ϵ^2) that

$$\mathrm{H}_{\epsilon/3}^{(j)} \leqq \tfrac{1}{4}\cdot 5\sum_{k=1}^{\infty} 2^k\cdot 64\rho_{jk}^2/(\tfrac{1}{9}\epsilon^2) = (720/\epsilon^2)\sum_{k=1}^{\infty} 2^k\rho_{jk}^2.$$

Using this estimate in (24),

$$H_\epsilon \leqq H(U) + (720/\epsilon^2)\sum_{k=1}^{\infty} 2^k\left(\sum_{j=1}^{\infty} \mu(A_j)\rho_{jk}^2\right).$$

In A_j, $\|y_k\| \geqq \rho_{jk}$. Hence

$$\sum_{j=1}^{\infty} \mu(A_j)\rho_{jk}^2 \leqq E(\|y_k\|^2) = a_k^2.$$

Thus

$$H_\epsilon \leqq H(U) + 720\, S/\epsilon^2.$$

We have from (23)

$$H(U) \leqq 4e[\log (9eS/\epsilon^2)]^2 \leqq 144S/\epsilon^2,$$

for $2\epsilon^2 \leqq 9S$. By (21), if $\epsilon^2 \leqq \sum n\sigma_n^2$, then certainly $\epsilon^2 \leqq \frac{1}{2}9S$, and $H_\epsilon \leqq 864S/\epsilon^2 \leqq 1728\sum n\sigma_n^2/\epsilon^2$. Theorem 9 is proved.

REMARK. It follows from (22) that for large ϵ the entropy has a bound of the form

$$\text{Const } (\sum n\sigma_n^2/\epsilon^2) \log (\epsilon^2/\sum n\sigma_n^2).$$

The following theorem shows that Theorem 9 is the best possible, in the sense that if $\sum n\sigma_n^2 = \infty$, there is a process under which $L_2[0, 1]$ has infinite ϵ-entropy. The construction can be shown to have the property that with probability 1, only finitely many ξ_n are nonzero. Thus, the sample functions can be made say entire functions with probability 1. We need one last lemma.

LEMMA 9. *Let $S(x, \epsilon)$ denote the closed spherical neighborhood of radius ϵ about a point x of the probabilistic metric space X. Then*

$$H_\epsilon(X) \geqq E(\log \{1/\mu[S(x, \epsilon)]\}).$$

PROOF. For any ϵ-partition U, let $p(x)$ denote the probability of the set of U containing x. Then

$$p(x) \leqq \mu\{S(x, \epsilon)\},$$

hence

$$H(U) = E(\log \{1/p(x)\}) \geqq E(\log \{1/\mu[S(x, \epsilon)]\}).$$

The conclusion follows immediately, since $H_\epsilon(X)$ is inf $H(U)$ over all ϵ-partitions. Lemma 9 is proved, and we come to Theorem 10, the last of this paper.

THEOREM 10. *If $\sum n\sigma_n^2 = \infty$, there is mean-continuous stochastic process on $[0, 1]$ with eigenvalues $\{\sigma_n^2\}$ which has infinite ϵ-entropy for every $\epsilon > 0$.*

PROOF. Let $\alpha_k = \sigma_{2^k}$, $k \geqq 0$. If we define

$$\sigma'_n = \alpha_k, \qquad 2^{k-1} < n \leqq 2^k, \quad k \geqq 0,$$

we have $\sigma'_n \leqq \sigma_n$ for all n. An infinite-entropy process will be set up for these eigenvalues. The process for the original eigenvalues can be obtained by suitably expanding the scales of the random variables involved, which can only increase the entropy. Since

$$\sum_{n=2^k+1}^{2^{k+1}} n\sigma_n^2 \leqq 2^{2k+1}\alpha_k^2 \leqq 8\sum_{n=2^{k-2}+1}^{2^k} n(\sigma'_n)^2,$$

the new sequence of eigenvalues satisfies the hypothesis, which can also be stated as $\sum_{k=1}^{\infty} 2^{2k}\alpha_k^2 = \infty$. Each eigenvalue α_k^2 has multiplicity ν_k, where $\nu_0 = 1$, $\nu_k = 2^{k-1}$, $k \geqq 1$.

The stochastic process will be set up in the form

$$x(t) = \sum_{n=1}^{\infty} \xi_n\phi_n(t),$$

where $\{\phi_n(t)\}$ is any orthonormal sequence of bounded continuous functions on $[0, 1]$, and $\{\xi_n\}$ is an orthogonal sequence of random variables with $E(\xi_n) = 0$, $E(\xi_n^2) = (\sigma'_n)^2$. The $\{\xi_n\}$ are defined below.

For each α_k, there is to be a set of ν_k random variables ξ_n with variance α_k^2. These ξ_n's form a ν_k-vector y_k which is to be independent of all other ξ_n's. This y_k has a distribution composed of two parts. The first is a uniform distribution on the unit sphere in ν_k-dimensional space, with probability $p_k = \nu_k\alpha_k^2$ (we assume without loss of generality that $\nu_k\alpha_k^2 < 1$). For the second part, $y_k = 0$ with probability $1 - p_k$. From the symmetry of this distribution, it follows easily that for ξ_n a component of y_k, $E(\xi_n) = 0$, $E(\xi_n^2) = \alpha_k^2$, and the ξ_n's are uncorrelated.

It will be shown that this process has infinite ϵ-entropy for $\epsilon < \frac{1}{2}$, by showing that the lower bound given by Lemma 9 is infinite.

Let a be fixed between 2ϵ and 1. By Lemma 6, there are at least a^{-m+1} points with mutual distances $\geqq a$ on the unit sphere in m dimensions. Hence, the probability that y_k lies in an ϵ-neighborhood on the unit sphere in ν_k dimensions is less than $a^{\nu_k-1}p_k$. Let $x(t)$ be such that $y_{k_1}, y_{k_2}, \cdots, y_{k_l}$ are non-zero, and the rest of the y_k's are zero. Then

$$\mu[S(x, \epsilon)] < \textstyle\prod_{k=0}^{\infty} (1 - p_k)\prod_{j=1}^{l} a^{\nu_{k_j}-1}p_{k_j}/(1 - p_{k_j});$$

$x(t)$ has this property with probability $\prod_{k=0}^{\infty} (1 - p_k)\prod_{j=1}^{l} p_{k_j}/(1 - p_{k_j})$.

Hence,

$$\begin{aligned} H_\epsilon(X) &\geqq E(\log \{1/\mu[S(x, e)]\}) \\ &\geqq \textstyle\sum_{\{k_j\}} \prod_{k=0}^{\infty} (1 - p_k)\prod_{j=1}^{l} [p_{k_j}/(1 - p_{k_j})] \{\log [1/\prod_{k=0}^{\infty} (1 - p_k)] \\ &\qquad + \textstyle\sum_{j=1}^{l} [\log (1 - p_{k_j})/p_{k_j} + (\nu_{kj} v_j - 1) \log (1/a)]\}. \end{aligned}$$

Simplifying this expression, we find

$$H_\epsilon \geqq \log [1/\textstyle\prod_{k=0}^{\infty} (1 - p_k)] + \sum_{k=0}^{\infty} p_k\{\log [(1 - p_k)/p_k] + (\nu_k - 1) \log (1/a)\}.$$

This is infinite because of the last term in the braces, for

$$\textstyle\sum_{k=1}^{\infty} (\nu_k - 1)p_k = \sum_{k=1}^{\infty} 2^{2k-2}\alpha_k^2 - 1$$

is infinite.

The process used above was only shown to have infinite ϵ-entropy for $\epsilon < \frac{1}{2}$. However, a similar construction gives a process which can be shown by the same method to have infinite ϵ-entropy for any $\epsilon > 0$. This can be accomplished by taking the sphere used in the distribution of y_k to have radius R_k, with $\lim_{k\to\infty} R_k = \infty$. Then let $p_k = \nu_k\alpha_k^2/R_k^2$, and infinite entropy follows if the R_k increase slowly enough that

$$\textstyle\sum_{k=1}^{\infty} (2^{2k}\alpha_k^2/R_k^2) = \infty.$$

Since such R_k always exist if $\sum 2^{2k}\alpha_k^2$ diverges, Theorem 10 is proved.

Theorem 10 can be used as in [8] to show for example that $\sum n\sigma_n^2 < \infty$ for processes on [0, 1] whose covariance function $R(s, t)$ satisfies $|2R(s, t) - R(s, s) - R(t, t)| = O(|s - t|^{1+\delta})$ as $|s - t| \to 0, \delta > 0$.

Remark. It can be shown that there exist mean-continuous stochastic processes on the unit interval (Gaussian in fact; see [6]) with finite ϵ-entropy for every positive ϵ, but such that H_ϵ is an arbitrarily rapidly increasing function as $\epsilon \to 0$. Of course, for these processes, $\sum n\sigma_n^2$ diverges. Examples can also be given of processes with finite ϵ-entropy for some ϵ but not others (these can be taken to have the support of μ be a bounded set in L_2).

We note that in [6] it is shown that the Wiener process on [0, 1] has ϵ-entropy H_ϵ with $1/2\epsilon^2 \lesssim H_\epsilon \leqq 1/\epsilon^2$. For this process, $\sigma_n^2 = 1/\pi^2(n - \frac{1}{2})^2$, $n \geqq 1$, and so $\sum n\sigma_n^2$ diverges, as it must by Theorem 9. By Theorem 10, however, there exists a non-Gaussian process with the very same eigenvalues whose ϵ-entropy is infinite for every $\epsilon > 0$.

As we have remarked previously, [6] shows that for any eigenvalues σ_n^2 with $\sum \sigma_n^2$ finite, no matter how slowly decreasing, there exists a process with those eigenvalues with finite ϵ-entropy for every positive ϵ. Namely, any Gaussian process with those eigenvalues will do. In other words, no given slow rate of decay of the eigenvalues guarantees infinite ϵ-entropy.

REFERENCES

[1] Feinstein, Amiel (1958). *Foundations of Information Theory.* McGraw-Hill, New York.
[2] Hausdorff, Felix (1957). *Set Theory.* Chelsea, New York.
[3] Loève, Michel (1955). *Probability Theory—Foundations. Random Sequences.* van Nostrand, New York.
[4] Neveu, Jacques (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.
[5] Posner, Edward C., Rodemich, Eugene R. and Rumsey, Howard, Jr. (1967). Epsilon entropy and data compression. In preparation.
[6] Posner, Edward C., Rodemich, Eugene R. and Rumsey, Howard, Jr. (1967). Epsilon entropy of Gaussian process. In preparation.
[7] Prokhorov, Yu. V. (1956). Convergence of random processes and limit theorems in probability theory (translation). *Theor. Prob. Appl.* **1** 157–214.
[8] Rodemich, Eugene R. (1967). Epsilon entropy of continuous processes. In preparation.
[9] Shannon, Claude E. (1948). A mathematical theory of communication. *Bell Systems Tech. J.* **27** 379–423.