ON THE DIMENSION AND ENTROPY OF PROBABILITY DISTRIBUTIONS

By

A. RÉNYI (Budapest), member of the Academy

Let η be a discrete random variable, taking on the different values y_k with the probabilities q_k $(k=1,2,\ldots)$, i. e.

$$\mathbf{P}(\eta=y_k)=q_k \quad \text{where} \quad q_k\geqq 0 \quad (k=1,2,\ldots) \quad \text{and} \quad \sum_{k=1}^{\infty} q_k=1.$$

(Here and in what follows $\mathbf{P}(\ldots)$ denotes the probability of the event in the brackets.) The entropy of η (which may also be called the entropy of the probability distribution of η) as defined by SHANNON, will be denoted by $\mathbf{H}_0(\eta)$; i. e. we put[1]

$$\mathbf{H}_0(\eta)=\sum_{k=1}^{\infty} q_k \log\frac{1}{q_k}, \tag{1}$$

provided that the series on the right of (1) converges. (If this series is divergent, we say that the entropy of η does not exist.)

Let now ξ be an arbitrary real-valued random variable, having the distribution function $F(x)$. The dimension and entropy of ξ have been defined in [1] as follows:

Put $\xi_n=\frac{1}{n}[n\xi]$ where $[x]$ denotes the integral part of x. If $\mathbf{H}_0([\xi])$ exists and if for $n\to+\infty$ the following asymptotic formula holds:[2]

$$\mathbf{H}_0(\xi_n)=d\log n+h+o(1) \tag{2}$$

(where $o(1)$ means a remainder term tending to 0 for $n\to+\infty$), then we shall say that the dimension $\mathbf{d}(\xi)$ of ξ (of the distribution of ξ) is equal to d and the d-dimensional entropy $\mathbf{H}_d(\xi)$ of ξ is equal to h; thus we put

$$\mathbf{d}(\xi)=\lim_{n\to+\infty}\frac{\mathbf{H}_0(\xi_n)}{\log n}, \tag{3}$$

provided this limit exists; otherwise the dimension of ξ is not defined. If the limit (3) exists, we put

$$\mathbf{H}_{\mathbf{d}(\xi)}(\xi)=\lim_{n\to+\infty}(\mathbf{H}_0(\xi_n)-\mathbf{d}(\xi)\log n), \tag{4}$$

[1] Throughout the paper log denotes the logarithm with respect to the base 2.

[2] Clearly, $\mathbf{H}_0(\xi_n)=\mathbf{H}_0([n\xi])$; we prefer to write $\mathbf{H}_0(\xi_n)$ because the distribution of ξ_n tends to that of ξ for $n\to+\infty$ and this shows why definition (2) is natural.

provided that the limit on the right of (4) exists; otherwise the entropy of ξ is not defined. In case the limit (3) does not exist, we may consider the lower and upper limits

$$\underline{\mathbf{d}}(\xi) = \varliminf_{n \to +\infty} \frac{\mathbf{H}_0(\xi_n)}{\log n} \tag{5}$$

and

$$\overline{\mathbf{d}}(\xi) = \varlimsup_{n \to +\infty} \frac{\mathbf{H}_0(\xi_n)}{\log n}, \tag{6}$$

and call $\underline{\mathbf{d}}(\xi)$ and $\overline{\mathbf{d}}(\xi)$ the lower and upper dimensions of ξ, respectively. Thus ξ has a definite dimension if and only if $\underline{\mathbf{d}}(\xi) = \overline{\mathbf{d}}(\xi)$.

The lower and upper dimensions, respectively, are not always defined either; as a matter of fact, if the quantities $\mathbf{H}_0(\xi_n)$ are all infinitely large, then $\underline{\mathbf{d}}(\xi)$ and $\overline{\mathbf{d}}(\xi)$ are not defined. If, however, $\mathbf{H}_0([\xi])$ is finite, then $\underline{\mathbf{d}}(\xi)$ and $\overline{\mathbf{d}}(\xi)$ always exist and we have

$$0 \leqq \underline{\mathbf{d}}(\xi) \leqq \overline{\mathbf{d}}(\xi) \leqq 1. \tag{7}$$

This can be shown as follows: Put

$$p_{nk} = \mathbf{P}\left(\frac{k}{n} \leqq \xi < \frac{k+1}{n}\right) \qquad (k = 0, \pm 1, \pm 2, \ldots). \tag{8}$$

We need the well-known inequality[3]

$$\frac{\sum_k a_k \log b_k}{\sum_k a_k} \leqq \log \frac{\sum_k a_k b_k}{\sum_k a_k} \tag{9}$$

valid for any finite sequences of non-negative numbers a_k and b_k such that $\sum_k a_k > 0$. It follows from (9) that

$$\sum_{k=ln}^{(l+1)n-1} p_{nk} \log \frac{1}{p_{nk}} \leqq p_{1l} \log \frac{n}{p_{1l}} \quad \text{for} \quad l = 0, \pm 1, \pm 2, \ldots \tag{10}$$

and thus, summing (10) for l, we obtain

$$0 \leqq \mathbf{H}_0(\xi_n) \leqq \mathbf{H}_0(\xi_1) + \log n, \tag{11}$$

which implies that if $\mathbf{H}_0(\xi_1)$ is finite, then $\mathbf{H}_0(\xi_n)$ is finite for all n and (7) holds. On the other hand, using the inequality (see [1])

$$\mathbf{H}_0(g(\xi)) \leqq \mathbf{H}_0(\xi), \tag{12}$$

valid for any discrete random variable ξ and for any function $g(\xi)$, as $\xi_1 = [\xi_k]$

[3] (9) is nothing else than Jensen's inequality applied to the concave function $\log x$; it may also be considered as the generalized form of the inequality between the geometric and arithmetic means. See e. g. G. Pólya and G. Szegő, *Aufgaben und Lehrsätze aus der Analysis. I* (Berlin, 1955), p. 53.

for any $k \geqq 1$ we have

$$\mathbf{H}_0(\xi_1) \leqq \mathbf{H}_0(\xi_k) \qquad (k=1,2,\ldots), \tag{13}$$

and thus if $\mathbf{H}_0(\xi_k)$ is finite for some k, then $\mathbf{H}_0(\xi_1)$ is finite and thus $\mathbf{H}_0(\xi_n)$ is finite for every n.

Thus we may say that every random variable ξ for which $\mathbf{H}_0([\xi])$ is finite, has upper and lower dimensions for which (7) holds. This is the case e. g. for every bounded random variable.

It is not difficult to give an example[4] of a random variable ξ for which $\mathbf{H}_0([\xi])$ is finite and for which the upper and lower dimensions are different.

Let us define the measure ν on the Borel subsets of the interval $I=[0,1)$ as follows: Let $a_1, a_2, \ldots$ be a sequence of positive integers which will be specified later. Let us divide the interval I in 2^{a_1} subintervals of length $\frac{1}{2^{a_1}}$ and let us attribute to the intervals $I_{10}=\left[0, \frac{1}{2^{a_1}}\right)$ and $I_{11}=\left[1-\frac{1}{2^{a_1}}, 1\right)$ the measure $\frac{1}{2}$ and to the other subintervals $\left[\frac{k}{2^{a_1}}, \frac{k+1}{2^{a_1}}\right)$ $(1 \leqq k \leqq 2^{a_1}-2)$ (if there are such subintervals, i. e. if $a_1>1$) the measure 0. Let us divide the intervals I_{10} and I_{11} into 2^{a_2} equal subintervals and attribute to the first and last ones which we denote by I_{20}, I_{21} and I_{22}, I_{23}, respectively, the measure $\frac{1}{4}$ and (if $a_2>1$) to the others the measure 0. Let us divide each of the intervals I_{2j} $(j=0,1,2,3)$ into 2^{a_3} equal subintervals and attribute to the first and last subintervals of each of these four intervals the measure $\frac{1}{8}$ and to the others (if there are any) the measure 0. Continuing this procedure ad infinitum, we have defined the measure ν for every subinterval of the form

$$\left[\frac{k}{2^{s_n}}, \frac{k+1}{2^{s_n}}\right) \qquad (k=0,1,\ldots,s_n-1)$$

where

$$s_n = a_1 + a_2 + \cdots + a_n \qquad (n=1,2,\ldots).$$

This measure ν can be extended in the usual way to a measure defined on all Borel subsets of the interval I; the measure ν satisfies evidently the condition $\nu(I)=1$, i. e. ν is a probability measure.

Let now ξ_ν be a random variable the distribution function $F(x)$ of which is equal to $\nu(I_x)$ where I_x denotes the interval $[0,x)$. (Note that in case $a_n=1$ for all n, $F(x)\equiv x$ for $0 \leqq x \leqq 1$, i. e. ν is the ordinary Lebesgue measure and ξ_ν is uniformly distributed in the interval $[0, 1)$.) Clearly, putting

4 A similar example has been constructed by T. Kővári (oral communication).

$\xi_n = \frac{1}{n}[n\xi]$, $\xi_{2^{s_n}}$ takes on 2^n different values each with the probability $\frac{1}{2^n}$, thus

$$\frac{\mathbf{H}_0(\xi_{2^{s_n}})}{\log 2^{s_n}} = \frac{n}{s_n}.$$

Thus, if we choose the sequence a_n in such a manner that the limit $\lim_{n\to+\infty} \frac{n}{s_n}$ does not exist, the lower and upper dimensions of ξ will be different, namely we shall have

$$\underline{\mathbf{d}}(\xi) \leqq \varliminf_{n\to+\infty} \frac{n}{s_n}$$

and

$$\overline{\mathbf{d}}(\xi) \geqq \varlimsup_{n\to+\infty} \frac{n}{s_n}.$$

For instance, if $a_k = 1$ for $2^{2r} \leqq k < 2^{2r+1}$ and $a_k = 2$ for $2^{2r+1} \leqq k < 2^{2r+2}$ $(r = 0, 1, 2, \ldots)$, then

$$\underline{\mathbf{d}}(\xi) \leqq \frac{3}{5} \quad \text{and} \quad \overline{\mathbf{d}}(\xi) \geqq \frac{3}{4}.$$

Evidently, by choosing the sequence a_n in another appropriate manner, we can reach that $\underline{\mathbf{d}}(\xi)$ and $\overline{\mathbf{d}}(\xi)$ shall have any prescribed values satisfying (7).

To prove that our definition is not contradictory we have to show that if the distribution of ξ is of the discrete type, ξ taking on the values x_k with the corresponding probabilities p_k ($x_j \neq x_k$ for $j \neq k$), then $\mathbf{d}(\xi) = 0$ and

$$\mathbf{H}_0(\xi) = \sum_k p_k \log \frac{1}{p_k},$$

provided that the series on the right converges. This can easily be shown as follows: Clearly, the values of ξ_n can be arranged in such a manner that denoting by $p_k^{(n)}$ $(k = 1, 2, \ldots)$ the corresponding probabilities we have for every fixed k

$$\lim_{n\to+\infty} p_k^{(n)} = p_k.$$

As by (12) we have, in case the series $\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$ converges,

$$\mathbf{H}_0(\xi_n) = \sum_{k=1}^{\infty} p_k^{(n)} \log \frac{1}{p_k^{(n)}} \leqq \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k},$$

it follows that for any N

$$\sum_{k=1}^{N} p_k \log \frac{1}{p_k} \leqq \varliminf_{n\to+\infty} \mathbf{H}_0(\xi_n) \leqq \varlimsup_{n\to+\infty} \mathbf{H}_0(\xi_n) \leqq \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$$

and this implies

$$\lim_{n\to-\infty} \mathbf{H}_0(\xi_n) = \mathbf{H}_0(\xi).$$

If the distribution of ξ is of the discrete type and $\mathbf{H}_0([\xi])$ is finite but $\mathbf{H}_0(\xi) = +\infty$, then still we have $\mathbf{d}(\xi) = 0$. This can be shown as follows: We may arrange again the values of ξ_n in such a manner that if $p_k^{(n)}$ denote the corresponding probabilities, we have for $n \geqq n_0(\varepsilon)$

$$|p_k^{(n)} - p_k| \leqq \frac{\varepsilon}{N} \quad \text{for} \quad k = 1, 2, \ldots, N$$

where N is chosen so that

$$\sum_{N+1}^{\infty} p_k < \varepsilon$$

should hold. It follows that for $n \geqq n_0(\varepsilon)$ $\sum_{k=N+1}^{\infty} p_k^{(n)} \leqq 2\varepsilon$. Clearly, $\sum_{k=1}^{N} p_k^{(n)} \log \frac{1}{p_k^{(n)}}$ remains bounded if $n \to +\infty$, further by (10)

$$\sum_{k=N+1}^{\infty} p_k^{(n)} \log \frac{1}{p_k^{(n)}} \leqq \mathbf{H}_0([\xi]) + 2\varepsilon \log n + O(1)$$

which implies that

$$\overline{\lim_{n\to+\infty}} \frac{\mathbf{H}_0(\xi_n)}{\log n} \leqq 2\varepsilon;$$

as $\varepsilon > 0$ is arbitrary, it follows that $\mathbf{d}(\xi) = 0$.

Now we turn to the investigation of absolutely continuous distributions and prove the following

THEOREM 1. *If ξ is a random variable having an absolutely continuous distribution with the density function $f(x)$ and if $\mathbf{H}_0([\xi])$ is finite, then $\mathbf{d}(\xi) = 1$ and we have*

$$\mathbf{H}_1(\xi) = \int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)}\, dx, \tag{14}$$

provided that the integral on the right of (14) *exists.*

REMARK. It should be mentioned that the existence of the integral on the right of (14) does not make unnecessary the condition that $\mathbf{H}_0([\xi])$ should be finite, as there exist distributions for which $\int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)}\, dx$ exists, but nevertheless $\mathbf{H}_0([\xi])$ is not finite. For instance, if $f(x)=\frac{C}{\log^2 n}$ for $n \leqq x \leqq n+\frac{1}{n}$ $(n=2,3,\ldots)$ where $C=\left(\sum_{n=2}^{\infty} \frac{1}{n \log^2 n}\right)^{-1}$ and $f(x)=0$ for all other values of x, then clearly

$$\int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)}\, dx = C \sum_{n=2}^{\infty} \frac{2 \log\log n + \log \frac{1}{C}}{n \log^2 n}$$

exists, but as

$$p_{1l} = \mathbf{P}(l \leqq \xi < l+1) = \frac{C}{l \log^2 l} \qquad (l=2,3,\ldots),$$

the series

$$\sum_l p_{1l} \log \frac{1}{p_{1l}}$$

is divergent and thus $\mathbf{H}_0([\xi])$ is not finite. Thus Theorem 1 does not justify the use of the value $\int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)}\, dx$ as the entropy of the distribution having the density function $f(x)$ always when it exists, but only in the case when $\mathbf{H}_0([\xi])$ is finite.

PROOF OF THEOREM 1. Let us first prove that $\mathbf{d}(\xi)=1$. Let us put

$$\varphi_n(x) = n \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(t)\, dt = n p_{nk} \quad \text{for} \quad \frac{k}{n} \leqq x < \frac{k+1}{n}. \tag{15}$$

We have clearly

$$\int_{-\infty}^{+\infty} \varphi_n(x)\, dx = \int_{-\infty}^{+\infty} f(t)\, dt = 1. \tag{16}$$

Then evidently

$$\sum_{k=-\infty}^{+\infty} p_{nk} \log \frac{1}{n p_{nk}} = \int_{-\infty}^{+\infty} f(x) \log \frac{1}{\varphi_n(x)}\, dx. \tag{17}$$

Let us first consider the case in which

$$\mathfrak{I} = \int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)}\, dx \tag{18}$$

exists. We use the well-known inequality [5]

(19) $$\frac{\int_E g(x)\log h(x)\,dx}{\int_E g(x)\,dx} \leqq \log \frac{\int_E g(x)h(x)\,dx}{\int_E g(x)\,dx}$$

valid for any non-negative measurable functions $g(x)$ and $h(x)$ and for any measurable set E for which the integrals in question exist and $\int_E g(x)\,dx > 0$. We obtain by (19) and (16)

(20) $$\mathfrak{I} - \sum_{k=-\infty}^{+\infty} p_{nk} \log \frac{1}{n p_{nk}} = \int_{-\infty}^{+\infty} f(x) \log \frac{\varphi_n(x)}{f(x)}\,dx \leqq \log \int_{-\infty}^{+\infty} \varphi_n(x)\,dx = 0$$

and thus

(21) $$\mathfrak{I} + \log n \leqq \mathbf{H}_0(\xi_n).$$

It follows from (11) and (21) that

$$\lim_{n\to+\infty} \frac{\mathbf{H}_0(\xi_n)}{\log n} = 1.$$

In the general case put for every $A>0$

(22) $$f_A(x) = \begin{cases} f(x) & \text{if } f(x) \leqq A, \\ 0 & \text{if } f(x) > A. \end{cases}$$

Evidently we have, putting

(23) $$\int_{-\infty}^{+\infty} f_A(x)\,dx = S(A),$$

(24) $$\lim_{A\to+\infty} S(A) = 1.$$

Let us put

(25) $$p_{nk}(A) = \int_{\frac{k}{n}}^{\frac{k+1}{n}} f_A(x)\,dx$$

and

(26) $$\varphi_n(A,x) = n p_{nk}(A) \quad \text{for} \quad \frac{k}{n} \leqq x < \frac{k+1}{n}.$$

Then we have

(27) $$\int_{-\infty}^{+\infty} \varphi_n(A,x)\,dx = S(A).$$

[5] See e. g. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities* (Cambridge, 1934), p. 167.

Let us choose n so large that for every k we have

$$p_{nk} \leqq \frac{1}{e}. \tag{28}$$

This is clearly possible. Taking into account that $x \log \frac{1}{x}$ is increasing for $0 \leqq x \leqq \frac{1}{e}$ and that $p_{nk}(A) \leqq p_{nk}$, the series $\sum_{k=-\infty}^{+\infty} p_{nk}(A) \log \frac{1}{p_{nk}(A)n}$ converges and we have

$$\mathbf{H}_0(\xi_n) - S(A) \log n \geqq \sum p_{nk}(A) \log \frac{1}{np_{nk}(A)} = \int_{-\infty}^{+\infty} f_A(x) \log \frac{1}{\varphi_n(A, x)} dx. \tag{29}$$

According to the inequality (19), we have[6]

$$\int_{-\infty}^{+\infty} f_A(x) \log \frac{1}{\varphi_n(A, x)} dx \geqq \int_{-\infty}^{+\infty} f_A(x) \log \frac{1}{f_A(x)} dx \geqq S(A) \log \frac{1}{A}. \tag{30}$$

It follows that

$$\frac{\mathbf{H}_0(\xi_n)}{\log n} \geqq S(A) + \frac{\int_{-\infty}^{+\infty} f_A(x) \log \frac{1}{f_A(x)} dx}{\log n} \tag{31}$$

and thus

$$\underline{\mathbf{d}}(\xi) \geqq S(A).$$

As A can be choosen arbitrarily large, this and (24) imply that $\underline{\mathbf{d}}(\xi) = 1$ and thus $\mathbf{d}(\xi) = 1$.

Let us now turn to the proof of the second part of Theorem 1.

We distinguish here also two cases, according to whether $f(x)$ is bounded or not. If $f(x)$ is bounded, $f(x) \leqq B$ say, let us put

$$g_n(x) = np_{nk} \log \frac{1}{np_{nk}} \quad \text{for} \quad \frac{k}{n} \leqq x < \frac{k+1}{n} \qquad (k = 0, \pm 1, \ldots). \tag{32}$$

Then we have

$$np_{nk} \leqq B.$$

Thus it follows that $g_n(x)$ is bounded

$$|g_n(x)| \leqq \max\left(\frac{\log e}{e}, B|\log B|\right). \tag{33}$$

[6] (30) implies that the integral $\int_{-\infty}^{+\infty} f_A(x) \log \frac{1}{f_A(x)} dx$ is convergent.

On the other hand, if $F(x)$ denotes the distribution function of ξ, i. e. $F(x)=\int_{-\infty}^{x} f(t)\,dt$, then for $\frac{k}{n} \leqq x < \frac{k+1}{n}$

$$g_n(x)=\left(\frac{F\left(\frac{k+1}{n}\right)-F\left(\frac{k}{n}\right)}{\frac{1}{n}}\right)\log\frac{1}{\left(\frac{F\left(\frac{k+1}{n}\right)-F\left(\frac{k}{n}\right)}{\frac{1}{n}}\right)}. \tag{34}$$

As $F'(x)=f(x)$ almost everywhere, we have

$$\lim_{n\to+\infty} g_n(x)=f(x)\log\frac{1}{f(x)} \tag{35}$$

almost everywhere.

Now, clearly, for any integer $L>0$

$$\sum_{-Ln\leqq k<+Ln} p_{nk}\log\frac{1}{np_{nk}}=\int_{-L}^{+L} g_n(x)\,dx. \tag{36}$$

Thus by the theorem of Lebesgue it follows that

$$\lim_{n\to+\infty}\sum_{-Ln\leqq k<Ln} p_{nk}\log\frac{1}{np_{nk}}=\int_{-L}^{+L} f(x)\log\frac{1}{f(x)}\,dx. \tag{37}$$

On the other hand, using the inequality (9) we obtain

$$\sum_{k<-Ln \text{ or } Ln\leqq k} p_{nk}\log\frac{1}{np_{nk}} \leqq \sum_{l<-L \text{ or } l\geqq L} p_{1l}\log\frac{1}{p_{1l}} \tag{38}$$

and by inequality (19)

$$\sum_{k<-Ln \text{ or } k\geqq Ln} p_{nk}\log\frac{1}{np_{nk}}=\int_{|x|>L} f(x)\log\frac{1}{\varphi_n(x)}\,dx \geqq \int_{|x|>L} f(x)\log\frac{1}{f(x)}\,dx. \tag{39}$$

According to our suppositions for any $\varepsilon>0$, we can choose $L=L(\varepsilon)$ so large that

$$\left|\int_{|x|>L} f(x)\log\frac{1}{f(x)}\,dx\right|<\varepsilon \tag{40}$$

and

$$\sum_{l<-L \text{ or } l\geqq L} p_{1l}\log\frac{1}{p_{1l}}<\varepsilon. \tag{41}$$

Thus we obtain from (37)—(41)

$$\lim_{n\to+\infty}(\mathbf{H}_0(\xi_n)-\log n)=\int_{-\infty}^{+\infty}f(x)\log\frac{1}{f(x)}\,dx. \tag{42}$$

The general case can be reduced to the special case in which $f(x)$ is bounded as follows: Let us put

$$r_{nk}(A)=p_{nk}-p_{nk}(A), \tag{43}$$

then putting

$$R(A)=\sum_{k=-\infty}^{+\infty}r_{nk}(A)=1-S(A) \tag{44}$$

we have

$$\lim_{A\to+\infty}R(A)=0. \tag{45}$$

Now, clearly,

$$\sum_{k=-\infty}^{+\infty}p_{nk}\log\frac{1}{np_{nk}}=\int_{-\infty}^{+\infty}f(x)\log\frac{1}{\varphi_n(x)}\,dx\geqq\int_{-\infty}^{+\infty}f(x)\log\frac{1}{f(x)}\,dx, \tag{46}$$

thus it suffices to prove

$$\overline{\lim_{n\to+\infty}}\sum_{k=-\infty}^{+\infty}p_{nk}\log\frac{1}{np_{nk}}\leqq\int_{-\infty}^{+\infty}f(x)\log\frac{1}{f(x)}\,dx. \tag{47}$$

Now we have

$$\sum_{k=-\infty}^{+\infty}p_{nk}\log\frac{1}{np_{nk}}=\sum_{k=-\infty}^{+\infty}p_{nk}(A)\log\frac{1}{np_{nk}(A)}-$$
$$-\sum_{k=-\infty}^{+\infty}p_{nk}(A)\log\left(1+\frac{r_{nk}(A)}{p_{nk}(A)}\right)+\sum_{k=-\infty}^{+\infty}r_{nk}(A)\log\frac{1}{np_{nk}}. \tag{48}$$

As regards the first term, we obtain by what has been proved above

$$\lim_{n\to+\infty}\sum_{k=-\infty}^{+\infty}p_{nk}(A)\log\frac{1}{np_{nk}(A)}=\int_{-\infty}^{+\infty}f_A(x)\log\frac{1}{f_A(x)}\,dx. \tag{49}$$

Regarding the second term, using the inequality $\log(1+x)\leqq x\log e$ for $x>0$ we have

$$0\leqq\sum_{k=-\infty}^{+\infty}p_{nk}(A)\log\left(1+\frac{r_{nk}(A)}{p_{nk}(A)}\right)\leqq R(A)\log e. \tag{50}$$

As regards the third term, we have

$$\sum_{k=-\infty}^{+\infty}r_{nk}(A)\log\frac{1}{np_{nk}}\leqq\sum_{|k|>Ln}p_{nk}\log\frac{1}{np_{nk}}+\sum_{l=-L}^{L-1}\sum_{k=ln}^{(l+1)n-1}r_{nk}(A)\log\frac{1}{np_{nk}} \tag{51}$$

and thus by (9)

$$\sum_{k=-\infty}^{+\infty} r_{nk}(A)\log\frac{1}{np_{nk}} \leqq \sum_{|l|>L} p_{1l}\log\frac{1}{p_{1l}} + \tag{52}$$

$$+\sum_{l=-L}^{L-1}\left(\sum_{k=ln}^{(l+1)n-1} r_{nk}(A)\right)\log\frac{1}{\left(\sum_{k=ln}^{(l+1)n-1} r_{nk}(A)\right)}.$$

Thus, if $R(A) < \frac{1}{e}$, applying (9) again to the last term we have

$$\sum_{k=-\infty}^{+\infty} r_{nk}(A)\log\frac{1}{np_{nk}} \leqq \sum_{|l|>L} p_{1l}\log\frac{1}{p_{1l}} + R(A)\log\frac{2L}{R(A)}. \tag{53}$$

It follows from (48), (49), (50) and (53) that

$$\sum_{k=-\infty}^{+\infty} p_{nk}\log\frac{1}{np_{nk}} \leqq \int_{-\infty}^{+\infty} f_A(x)\log\frac{1}{f_A(x)}\,dx + \tag{54}$$

$$+\delta_n(A) + R(A)\left(\log 2L + \log\frac{1}{R(A)}\right) + \sum_{|l|>L} p_{1l}\log\frac{1}{p_{1l}}$$

where

$$\delta_n(A) = \sum_{k=-\infty}^{+\infty} p_{nk}(A)\log\frac{1}{np_{nk}(A)} - \int_{-\infty}^{+\infty} f_A(x)\log\frac{1}{f_A(x)}\,dx$$

and thus by (48) $\lim_{n\to+\infty}\delta_n(A) = 0.$

Clearly

$$\lim_{A\to+\infty}\int_{-\infty}^{+\infty} f_A(x)\log\frac{1}{f_A(x)}\,dx = \int_{-\infty}^{+\infty} f(x)\log\frac{1}{f(x)}\,dx. \tag{55}$$

Let us now choose L so large that

$$\sum_{|l|>L} p_{1l}\log\frac{1}{p_{1l}} < \varepsilon; \tag{56}$$

fixing L let us choose A so large that

$$R(A)\log\frac{2L}{R(A)} < \varepsilon, \tag{57}$$

further

$$\left|\int_{-\infty}^{+\infty} f_A(x)\log\frac{1}{f_A(x)}\,dx - \int_{-\infty}^{+\infty} f(x)\log\frac{1}{f(x)}\,dx\right| = \left|\int_{f(x)>A} f(x)\log\frac{1}{f(x)}\,dx\right| < \varepsilon, \tag{58}$$

and finally n_0 so large that

$$|\delta_n(A)| < \varepsilon \tag{59}$$

for $n \geqq n_0$ where n_0 depends only on ε; we obtain from (54), (56), (57), (58) and (59) that

$$\sum_{k=-\infty}^{+\infty} p_{nk} \log \frac{1}{n p_{nk}} \leqq \int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)} dx + 4\varepsilon. \tag{60}$$

As $\varepsilon > 0$ may be chosen arbitrarily small, we obtain (47). This proves Theorem 1.

Let us mention the following theorem which is a special case of Theorem 1:

THEOREM 2. *Let E be a bounded and measurable set on the real axis, having positive Lebesgue measure $\mu(E) > 0$. Let I_{nk} denote the interval $\left[\frac{k}{n}, \frac{k+1}{n}\right)$ $(k = 0, \pm 1, \pm 2, \ldots)$ and put*[7]

$$\mu_{nk} = \mu(E I_{nk}),$$

i. e. μ_{nk} denotes the Lebesgue measure of that part of the set E which lies in the interval I_{nk}. Then we have

$$\lim_{n \to +\infty} \sum_{k=-\infty}^{+\infty} \mu_{nk} \log \frac{1}{n \mu_{nk}} = 0.$$

REMARK. Clearly, in Theorem 2 it is irrelevant whether log denotes the natural logarithm or the logarithm with respect to the base 2.

PROOF OF THEOREM 2. Let us put

$$f(x) = \begin{cases} \dfrac{1}{\mu(E)} & \text{if } \ x \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Then, clearly, $f(x)$ is the density function of an absolutely continuous distribution, and thus if ξ is a random variable having $f(x)$ as its density function, then, as E is bounded, ξ is bounded and thus $\mathbf{H}_0([\xi])$ exists. Thus Theorem 1 can be applied; as

$$p_{nk} = \int_{k/n}^{k+1/n} f(x)\, dx = \frac{\mu_{nk}}{\mu(E)},$$

it follows that

$$\lim_{n \to +\infty} \sum_{k=-\infty}^{+\infty} \frac{\mu_{nk}}{\mu(E)} \log \frac{\mu(E)}{n \mu_{nk}} = \int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)} dx = \log \mu(E).$$

Thus the assertion of Theorem 2 follows.

[7] AB denotes the intersection of the sets A and B.

Let us consider now the case when the distribution of ξ is the mixture of a distribution of the discrete type and of an absolutely continuous distribution. For this case the following theorem holds:

THEOREM 3. *Let ξ be a random variable such that $\mathbf{H}_0([\xi])$ is finite and the distribution function $F(x)$ of ξ can be represented in the form*

$$F(x) = (1-d)F_0(x) + dF_1(x) \tag{61}$$

where $F_0(x)$ is a purely discontinuous distribution function, i. e. there exists a sequence x_k ($k = 1, 2, \ldots$; $x_j \neq x_k$ for $j \neq k$) and a corresponding sequence $p_k \geqq 0$ for which $\sum_{k=1}^{\infty} p_k = 1$, so that

$$F_0(x) = \sum_{x_k < x} p_k,$$

and $F_1(x)$ is an absolutely continuous distribution function, i. e.

$$F_1(x) = \int_{-\infty}^{x} f_1(t)\,dt,$$

further $0 < d < 1$; then the dimension of ξ is equal to the weight d of the absolutely continuous component, i. e.

$$\mathbf{d}(\xi) = d. \tag{62}$$

If, further, the series $\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$ and the integral $\int_{-\infty}^{+\infty} f_1(x) \log \frac{1}{f_1(x)}\,dx$ are both convergent, the d-dimensional entropy of ξ is given by

$$\mathbf{H}_d(\xi) = (1-d)\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k} + d \int_{-\infty}^{+\infty} f_1(x) \log \frac{1}{f_1(x)}\,dx + d \log \frac{1}{d} + (1-d) \log \frac{1}{1-d}. \tag{63}$$

REMARK. The special case of Theorem 3 in which the distribution $\{p_k\}$ consists only of a finite number of terms, further $f_1(x)$ is continuous except for a finite number of points and vanishes outside a finite interval, has been proved in [1].

PROOF OF THEOREM 3. Put

$$p_{nk}^{(0)} = F_0\left(\frac{k+1}{n}\right) - F_0\left(\frac{k}{n}\right) \tag{64}$$

and

$$p_{nk}^{(1)} = F_1\left(\frac{k+1}{n}\right) - F_1\left(\frac{k}{n}\right), \tag{65}$$

then we have

$$p_{nk}=F\left(\frac{k+1}{n}\right)-F\left(\frac{k}{n}\right)=(1-d)p_{nk}^{(0)}+dp_{nk}^{(1)}. \tag{66}$$

Now we have, in general, for $a>0$, $b>0$

$$0\leqq\left(a\log\frac{1}{a}+b\log\frac{1}{b}-(a+b)\log\frac{1}{a+b}\right)= \tag{67}$$
$$=(a+b)\left[\frac{a}{a+b}\log\frac{a+b}{a}+\frac{b}{a+b}\log\frac{a+b}{b}\right]\leqq a+b.$$

Thus, if ξ_0 and ξ_1 denote random variables having the distribution functions $F_0(x)$ and $F_1(x)$, respectively, further putting

$$\mathcal{H}(p)=p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p} \tag{68}$$

and

$$\lambda_{nk}=\frac{dp_{nk}^{(1)}}{p_{nk}}, \tag{69}$$

we obtain

$$0\leqq d\mathbf{H}_0([n\xi_1])+(1-d)\mathbf{H}_0([n\xi_0])-\mathbf{H}_0([n\xi])+\mathcal{H}(d)=\sum_{k=-\infty}^{+\infty}p_{nk}\mathcal{H}(\lambda_{nk}). \tag{70}$$

Let us choose an $\varepsilon>0$. Let $A_n(\varepsilon)$ denote the set of those indices k for which $\lambda_{nk}\leqq\varepsilon$ and $B_n(\varepsilon)$ the set of those indices k for which $1-\lambda_{nk}\leqq\varepsilon$. Let us choose a number $N_1=N_1(\varepsilon)$ such that $\sum_{l=N_1}^{\infty}p_l<\varepsilon^2$. Let us choose an $N_2=N_2(\varepsilon)$ such that $N_2>N_1$ and $\frac{1}{N_2}<|x_i-x_j|$ for $1\leqq i<j\leqq N_1$. Let $C_n(\varepsilon)$ denote the set of those values of k for which the interval $\left[\frac{k}{n},\frac{k+1}{n}\right)$ contains one of the values x_i $(i=1,2,\ldots,N_1)$. If $n\geqq N_2$, clearly any interval $\left[\frac{k}{n},\frac{k+1}{n}\right)$ can contain only at most one such x_i. Evidently, we can find an $N_3>N_2$ such that if $n\geqq N_3$ $k\in C_n(\varepsilon)$ and x_i $(1\leqq i\leqq N_1)$ is contained in the interval $\left[\frac{k}{n},\frac{k+1}{n}\right)$, then $p_{nk}^{(1)}\leqq\frac{\varepsilon(1-d)p_i}{d}$ and thus $\lambda_{nk}\leqq\varepsilon$. Thus, for $n\geqq 3$ the set $C_n(\varepsilon)$ is a subset of $A_n(\varepsilon)$. Let $D_n(\varepsilon)$ denote the set of those integers k which do not belong to the union of $A_n(\varepsilon)$ and $B_n(\varepsilon)$. Clearly, if $k\in D_n(\varepsilon)$, then

$$dp_{nk}^{(1)}<\left(\frac{1-\varepsilon}{\varepsilon}\right)p_{nk}^{(0)}(1-d)$$

and thus

$$p_{nk} \leqq \frac{1}{\varepsilon} p_{nk}^{(0)}.$$

As clearly $C_n(\varepsilon) \subset A_n(\varepsilon)$ implies

$$\sum_{k \in D_n(\varepsilon)} p_{nk}^{(0)} \leqq \varepsilon^2,$$

it follows, taking into account that $\mathcal{H}(p) \leqq 1$ for $0 \leqq p \leqq 1$ (with equality taking place if and only if $p = 1/2$), that

$$\sum_{k \in D_n(\varepsilon)} p_{nk} \mathcal{H}(\lambda_{nk}) \leqq \sum_{k \in D_n(\varepsilon)} p_{nk} \leqq \frac{1}{\varepsilon} \sum_{k \in D_n(\varepsilon)} p_{nk}^{(0)} \leqq \varepsilon. \tag{71}$$

On the other hand, as $\underset{0 \leqq \lambda \leqq \varepsilon}{\operatorname{Max}} \mathcal{H}(\lambda) = \underset{1-\varepsilon \leqq \lambda \leqq 1}{\operatorname{Max}} \mathcal{H}(\lambda) = \mathcal{H}(\varepsilon)$ for $0 < \varepsilon \leqq 1/2$, it follows that (denoting by $A \cup B$ the union of the sets A and B)

$$\sum_{k \in A_n(\varepsilon) \cup B_n(\varepsilon)} p_{nk} \mathcal{H}(\lambda_{nk}) \leqq \mathcal{H}(\varepsilon). \tag{72}$$

Thus we obtain from (70), (71) and (72) that for $n \geqq N_3 = N_3(\varepsilon)$

$$0 \leqq (1-d) \mathbf{H}_0([n\xi_0]) + d\mathbf{H}_0([n\xi_1]) - \mathbf{H}_0([n\xi]) + \mathcal{H}(d) \leqq \varepsilon + \mathcal{H}(\varepsilon). \tag{73}$$

As by Theorem 1

$$\lim_{n \to +\infty} \frac{\mathbf{H}_0([n\xi_1])}{\log n} = 1,$$

further

$$\lim_{n \to +\infty} \frac{\mathbf{H}_0([n\xi_0])}{\log n} = 0,$$

it follows

$$\lim_{n \to +\infty} \frac{\mathbf{H}_0([n\xi])}{\log n} = d. \tag{74}$$

As further in case $\sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$ is convergent, we have

$$\lim_{n \to +\infty} \mathbf{H}_0([n\xi_0]) = \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k},$$

and in case the integral $\int_{-\infty}^{+\infty} f_1(x) \log \frac{1}{f_1(x)} dx$ exists, we have by Theorem 1

$$\lim_{n \to +\infty} (\mathbf{H}_0([n\xi_1]) - \log n) = \int_{-\infty}^{+\infty} f_1(x) \log \frac{1}{f_1(x)} dx,$$

it follows from (73), as $\varepsilon>0$ may be chosen arbitrarily small, that

$$\lim_{n\to+\infty}(\mathbf{H}_0([n\xi])-d\log n)=(1-d)\sum_{k=1}^{\infty}p_k\log\frac{1}{p_k}+ \tag{75}$$
$$+d\int_{-\infty}^{+\infty}f_1(x)\log\frac{1}{f_1(x)}\,dx+\mathcal{H}(d).$$

Thus Theorem 3 is proved.

As it is well known, every probability distribution function $F(x)$ can be represented in the form

$$F(x)=pF_0(x)+qF_1(x)+rF_2(x) \tag{76}$$

where $F_0(x)$ is a purely discontinuous probability distribution function, $F_1(x)$ is an absolutely continuous probability distribution function and $F_2(x)$ is a continuous singular probability distribution function, $p\geqq 0$, $q\geqq 0$, $r\geqq 0$ and $p+q+r=1$. Thus Theorem 2 gives a full answer concerning the dimension and entropy of any real random variable which is such that its probability distribution function $F(x)$ has no singular component; i. e. is such that in (76) $r=0$. The question about the existence and (if it exists) the value of the dimension and entropy, resp., of a random variable having a singular distribution function seems to be rather intricate. (Note that the distribution function $F_\nu(x)$ mentioned above for which the lower and upper dimensions were shown to be different is of the singular type!)

The results obtained above can be generalized to s-dimensional vector-valued random variables or, by other words, to probability distributions in a Euclidean space E_s of s dimensions $(s=2,3,\ldots)$. To this purpose we have to extend the definition of the notions "dimension" and "entropy" to the case of a probability distribution in E_s. This can be done as follows: Let us define the symbol [ ] (integral part) for s-dimensional vectors as follows: if $\vec{x}$ is an s-dimensional vector with the real components $x_1, x_2, \ldots, x_s$, denote by $[\vec{x}]$ the vector having the components $[x_1], [x_2], \ldots, [x_s]$. If $\vec{\zeta}$ is a random s-dimensional vector, let us put $\vec{\zeta}_n=\frac{1}{n}[n\vec{\zeta}]$; $\vec{\zeta}$ is clearly a random variable having a distribution of the discrete type and thus its entropy $\mathbf{H}_0(\vec{\zeta}_n)$ is defined. We restrict ourselves to the case when $\mathbf{H}_0([\vec{\zeta}])$ is finite, in which case $\mathbf{H}_0(\vec{\zeta}_n)$ is finite for every $n=1,2,\ldots$.

The lower and upper "dimensions" $\underline{\mathbf{d}}(\vec{\zeta})$ and $\overline{\mathbf{d}}(\vec{\zeta})$ of ζ are defined by

$$\underline{\mathbf{d}}(\vec{\zeta})=\varliminf_{n\to+\infty}\frac{\mathbf{H}_0(\vec{\zeta}_n)}{\log n} \tag{77}$$

and

(78) $$\overline{\mathbf{d}}(\vec{\zeta}) = \overline{\lim_{n \to +\infty}} \frac{\mathbf{H}_0(\vec{\zeta}_n)}{\log n},$$

respectively. If $\underline{\mathbf{d}}(\vec{\zeta}) = \overline{\mathbf{d}}(\vec{\zeta})$, we call this value the dimension of $\vec{\zeta}$ and denote it by $\mathbf{d}(\vec{\zeta})$; i. e. we put

(79) $$\mathbf{d}(\vec{\zeta}) = \lim_{n \to +\infty} \frac{\mathbf{H}_0(\vec{\zeta}_n)}{\log n},$$

provided that the limit on the right of (79) exists. It is easy to show that

(80) $$0 \leqq \underline{\mathbf{d}}(\vec{\zeta}) \leqq \overline{\mathbf{d}}(\vec{\zeta}) \leqq s.$$

If $\mathbf{d}(\vec{\zeta}) = d$, we put

(81) $$\mathbf{H}_d(\vec{\zeta}) = \lim_{n \to +\infty} (\mathbf{H}_0(\vec{\zeta}_n) - d \log n),$$

provided that the limit on the right-hand side of (81) exists, and call $\mathbf{H}_d(\vec{\zeta})$ the d-dimensional entropy of $\vec{\zeta}$.

THEOREM 4. *Let $\vec{\zeta}$ be a random vector in E_s $(s = 2, 3, \ldots)$ and let us suppose that the distribution of $\vec{\zeta}$ is absolutely continuous with density function $f(\vec{x})$ where $\vec{x} = (x_1, \ldots, x_s)$; by other words, let us suppose for any Borel subset B of E_s*

(82) $$\mathbf{P}(\zeta \in B) = \int \cdots \int_B f(\vec{x})\, d\vec{x}.$$

Let us suppose, further, that $\mathbf{H}_0([\vec{\zeta}])$ is finite. Then we have

(83) $$\mathbf{d}(\vec{\zeta}) = s$$

and

(84) $$\mathbf{H}_s(\vec{\zeta}) = \int \cdots \int f(\vec{x}) \log \frac{1}{f(\vec{x})}\, d\vec{x},$$

provided that the integral on the right-hand side of (84) *exists.*

REMARK. Clearly, (83) can be expressed by saying that the geometrical (or topological) and information-theoretical concepts of dimension coincide for absolutely continuous probability distributions.

It can be shown that this coincidence is valid also for absolutely continuous probability distributions on sufficiently smooth p-dimensional manifolds lying in E_s with $p < s$; e. g. for an absolutely continuous probability distribution on the surface of a sphere in 3-space the dimension as defined above is equal to 2. It is a much more difficult question what is the topological background of a continuous distribution in E_s $(s = 1, 2, \ldots)$ having non-integral "dimension" $d < s$. It seems that an answer to this question can be given by using the notion of dimension introduced by HAUSDORFF.

Finally, we should like to show how the theory developed in the present paper can be generalized to random variables with values in an abstract metric space. For such variables A. N. KOLMOGOROV [2], [3] introduced the notion of ε-entropy. We shall here sketch a theory which is nothing else but a variant of the theory of KOLMOGOROV and which reduces to the theory of the present paper if the metric space in question is a compact subset of the Euclidean space of s dimensions $(s=1, 2, \ldots)$.[8]

Let X be a metric space with metric ϱ, which is completely bounded; by this we mean that for any $\varepsilon>0$ X can be subdivided into a finite number of non-overlapping sets each having a diameter $\leqq\varepsilon$. Let $N_X(\varepsilon)$ denote the minimal number of such sets (see [4]). Let $X_1(\varepsilon), \ldots, X_{n(\varepsilon)}(\varepsilon)$ be a system of non-overlapping sets whose union is equal to X and are such that each has the diameter $\leqq\varepsilon$. We do not require that the system $\{X_k(\varepsilon)\}$ be minimal, i. e. we do not suppose $n(\varepsilon)=N_X(\varepsilon)$, only that it should be "asymptotically minimal", in the sense that

$$\lim_{\varepsilon\to 0}\frac{\log n(\varepsilon)}{\log N_X(\varepsilon)}=1. \tag{86}$$

Let us suppose, further, that the sets $X_k(\varepsilon)$ belong to the least σ-algebra of subsets of X which contains all spheres, i. e. sets $S_a(r)$ of points x satisfying $\varrho(a, x)<r$ where $a\in X$ and $r>0$ are arbitrary; by other words, $X_k(\varepsilon)$ is a Borel subset of X.

Let ξ be a random variable with values in X, defined on a probability space $[\Omega, \mathcal{A}, \mathbf{P}]$, and if $Y\subset X$, denote by $\xi^{-1}(Y)$ the set of those $\omega\in\Omega$ for which $\xi(\omega)\in Y$. We consider only such random variables ξ for which for any $r>0$ and $a\in X$ the set $\xi^{-1}(S_a(r))$ is measurable, i. e. belongs to $\mathcal{A}$; let us put

$$p_k(\varepsilon)=\mathbf{P}(\xi^{-1}(X_k(\varepsilon))) \qquad (k=1, 2, \ldots, n(\varepsilon)) \tag{87}$$

[8] The difference between the definition of the ε-entropy given here and that given by KOLMOGOROV consists in that we introduce the notion of the "ε-entropy with respect to a given subdivision of the space X". This notion enables us to define ε-entropy without using the "amount of information"

$$\mathbf{I}(\xi, \eta)=\int \mathbf{P}_{\xi\eta}(dx, dy)\log\frac{\mathbf{P}_{\xi\eta}(dx, dy)}{\mathbf{P}_{\xi}(dx)\,\mathbf{P}_{\eta}(dy)} \tag{85}$$

on which KOLMOGOROV bases the definition of the ε-entropy. Thus the definition is somewhat simplified. Our ε-entropy depends, of course, on the subdivision considered; this dependence is, however, as we shall show in a particular case, rather weak. Besides that, the subdivision can always be chosen so as to simplify the evaluation of the ε-entropy, and this may be an advantage in certain cases.

and

$$\mathbf{H}(\varepsilon, \xi) = \sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon) \log \frac{1}{p_k(\varepsilon)}. \tag{88}$$

We shall call $\mathbf{H}(\varepsilon, \xi)$ *the ε-entropy of ξ with respect to the subdivision $\{X_k(\varepsilon)\}$ of X.* Clearly, by (9),

$$\mathbf{H}(\varepsilon, \xi) \leqq \log n(\varepsilon).$$

Let us define

$$\underline{\mathbf{D}}(\xi) = \varliminf_{\varepsilon \to 0} \frac{\mathbf{H}(\varepsilon, \xi)}{\log n(\varepsilon)} = \varliminf_{\varepsilon \to 0} \frac{\mathbf{H}(\varepsilon, \xi)}{\log N_X(\varepsilon)} \tag{89}$$

and

$$\overline{\mathbf{D}}(\xi) = \varlimsup_{\varepsilon \to 0} \frac{\mathbf{H}(\varepsilon, \xi)}{\log n(\varepsilon)} = \varlimsup_{\varepsilon \to 0} \frac{\mathbf{H}(\varepsilon, \xi)}{\log N_X(\varepsilon)}, \tag{90}$$

further put in case $\underline{\mathbf{D}}(\xi) = \overline{\mathbf{D}}(\xi)$

$$\mathbf{D}(\xi) = \lim_{\varepsilon \to 0} \frac{\mathbf{H}(\varepsilon, \xi)}{\log n(\varepsilon)} = \lim_{\varepsilon \to 0} \frac{\mathbf{H}(\varepsilon, \xi)}{\log N_X(\varepsilon)}. \tag{91}$$

If $\mathbf{D}(\xi)$ exists, the asymptotic expansion of $\mathbf{H}(\varepsilon, \xi)$ can be investigated further. Clearly, if X is the unit cube of the Euclidean space of s dimensions, and we choose for any $\varepsilon > 0$ the subdivision of X into cubes with sides equal to $\dfrac{1}{\left[\frac{\sqrt{s}}{\varepsilon}\right]+1}$ (i. e. with diameter $\dfrac{\sqrt{s}}{\left[\frac{\sqrt{s}}{\varepsilon}\right]+1} \leqq \varepsilon$), we have clearly for $\underline{\mathbf{D}}(\xi)$ and $\overline{\mathbf{D}}(\xi)$ defined by (89) and (90), resp.,

$$\underline{\mathbf{D}}(\xi) = \frac{\underline{\mathbf{d}}(\xi)}{s} \quad \text{and} \quad \overline{\mathbf{D}}(\xi) = \frac{\overline{\mathbf{d}}(\xi)}{s} \tag{92}$$

where $\underline{\mathbf{d}}(\xi)$ and $\overline{\mathbf{d}}(\xi)$ are the lower and upper dimensions of ξ as defined by (5) and (6), resp.

To show how weakly $\mathbf{H}(\varepsilon, \xi)$ depends on the choice of the set of subdivisions, let us consider the one-dimensional case in detail. Let X be the interval $[0, 1)$; let us choose for any $\varepsilon > 0$ a sequence $x_0(\varepsilon) = 0 < x_1(\varepsilon) < x_2(\varepsilon) < \cdots < x_{n(\varepsilon)}(\varepsilon) = 1$ such that $0 \leqq x_{k+1}(\varepsilon) - x_k(\varepsilon) \leqq \varepsilon$ $(k = 0, 1, \ldots, n(\varepsilon) - 1)$ and

$$\lim_{\varepsilon \to 0} \frac{\log n(\varepsilon)}{\log \frac{1}{\varepsilon}} = 1. \tag{93}$$

As clearly $\left[\frac{1}{\varepsilon}\right] \leqq N(\varepsilon) \leqq \left[\frac{1}{\varepsilon}\right] + 1$, if $X_k(\varepsilon)$ is the interval $[x_{k-1}(\varepsilon), x_k(\varepsilon))$, then this subdivision is admissible. Let ξ be an arbitrary random variable with values in the interval $(0, 1)$ whose distribution function $F(x)$ is absolutely continuous. Let $f(x)=F'(x)$ be the density function of ξ. Then we have, putting

$$d_k(\varepsilon)=x_k(\varepsilon)-x_{k-1}(\varepsilon) \qquad (k=1,2,\dots,n(\varepsilon))$$

and

$$p_k(\varepsilon)=F(x_k(\varepsilon))-F(x_{k-1}(\varepsilon)),$$

by (9)

$$\mathbf{H}(\varepsilon,\xi)=\sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon)\log\frac{1}{p_k(\varepsilon)}\leqq \log n(\varepsilon) \tag{94}$$

and thus by (93)

$$\overline{\mathbf{D}}(\xi)=\overline{\lim_{\varepsilon\to 0}}\frac{\mathbf{H}(\varepsilon,\xi)}{\log\frac{1}{\varepsilon}}\leqq 1. \tag{95}$$

On the other hand, as $d_k(\varepsilon)\leqq\varepsilon$, we have, putting

$$A(\varepsilon)=\sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon)\log\frac{d_k(\varepsilon)}{p_k(\varepsilon)}, \tag{96}$$

$$\mathbf{H}(\varepsilon,\xi)=\sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon)\log\frac{d_k(\varepsilon)}{p_k(\varepsilon)}+\sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon)\log\frac{1}{d_k(\varepsilon)}\geqq A(\varepsilon)+\log\frac{1}{\varepsilon}. \tag{97}$$

Now, clearly, by the same argument as used in proving Theorem 1, we obtain

$$\lim_{\varepsilon\to 0}\frac{A(\varepsilon)}{\log\frac{1}{\varepsilon}}=0. \tag{98}$$

Thus from (97) and (98) we obtain that

$$\underline{\mathbf{D}}(\xi)=\underline{\lim_{\varepsilon\to 0}}\frac{\mathbf{H}(\varepsilon,\xi)}{\log\frac{1}{\varepsilon}}\geqq 1. \tag{99}$$

(95) and (99) imply $\mathbf{D}(\xi)=1$.

Thus we have proved

THEOREM 5. *Let ξ be a real random variable. Let us suppose that the distribution of ξ is absolutely continuous with the density function $f(x)$, and suppose that the values of ξ are contained in the interval $[0,1)$, i. e. $\int_0^1 f(x)\,dx=1$. Let for any $\varepsilon>0$ be given a subdivision $0=x_0(\varepsilon)<x_1(\varepsilon)<\dots<x_{n(\varepsilon)}=1$ of the interval $[0,1)$ such that $x_k(\varepsilon)-x_{k-1}(\varepsilon)\leqq\varepsilon$ $(k=1,2,\dots,n(\varepsilon))$ and*

$$\lim_{\varepsilon\to 0}\frac{\log n(\varepsilon)}{\log\frac{1}{\varepsilon}}=1. \tag{100}$$

Let us put

(101) $$p_k(\varepsilon)=\int_{x_{k-1}(\varepsilon)}^{x_k(\varepsilon)} f(t)\,dt$$

and

(102) $$\mathbf{H}(\varepsilon,\xi)=\sum_{k=1}^{n} p_k(\varepsilon)\log\frac{1}{p_k(\varepsilon)}.$$

Then we have

(103) $$\lim_{\varepsilon\to 0}\frac{\mathbf{H}(\varepsilon,\xi)}{\log\frac{1}{\varepsilon}}=1.$$

Of course, if ξ is an arbitrary bounded variable, $|\xi|<B$, Theorem 5 can be applied to $\xi'=\frac{\xi+B}{2B}$ for which $0\leqq\xi'<1$. For the whole real axis the situation is somewhat different, because the whole real axis as a metric space with the metric $\varrho(x,y)=|x-y|$ is not completely bounded.

In the s-dimensional case $(s=2,3,\ldots)$ the situation is similar if we restrict ourselves to subdivisions of the unit cube of E_s into equal s-dimensional intervals, but we encounter some geometrical difficulties which do not present themselves in the one-dimensional case if we consider more general subdivisions.

As regards (for $s=1$) the difference $\mathbf{H}(\varepsilon,\xi)-\log\frac{1}{\varepsilon}$, it depends much more closely on the choice of the subdivision $\{x_k(\varepsilon)\}$.

In this direction we prove

THEOREM 6. *Let ξ be a real random variable. Let us suppose that the distribution of ξ is absolutely continuous with the density function $f(x)$; suppose that the values of ξ are contained in the interval* $[0,1)$ *and that the integral*

(104) $$\mathbf{H}_1(\xi)=\int_0^1 f(x)\log\frac{1}{f(x)}\,dx$$

exists. Let for any $\varepsilon>0$ be given a subdivision $0=x_0(\varepsilon)<x_1(\varepsilon)<\cdots<x_{n(\varepsilon)}(\varepsilon)=1$ of the interval $[0,1)$ *such that*

(105) $$\lim_{\varepsilon\to 0}\varepsilon n(\varepsilon)=1,$$

further putting

(106) $$d_k(\varepsilon)=x_k(\varepsilon)-x_{k-1}(\varepsilon)\qquad (k=1,2,\ldots,n(\varepsilon))$$

we have

(107) $$\frac{\varepsilon}{C}\leqq d_k(\varepsilon)\leqq\varepsilon$$

where $C>1$ is a positive constant not depending on ε. Let us define $p_k(\varepsilon)$ and $\mathbf{H}(\varepsilon, \xi)$ by (101) *and* (102), *respectively. Then we have*

$$\lim_{\varepsilon\to 0}\left(\mathbf{H}(\varepsilon,\xi)-\log\frac{1}{\varepsilon}\right)=\mathbf{H}_1(\xi). \tag{108}$$

PROOF OF THEOREM 6. We evidently have, putting

$$\Delta(\varepsilon)=\sum_{k=1}^{n(\varepsilon)} p_k(\varepsilon)\log\frac{\varepsilon}{d_k(\varepsilon)}, \tag{109}$$

$$\mathbf{H}(\varepsilon,\xi)=A(\varepsilon)+\Delta(\varepsilon)+\log\frac{1}{\varepsilon} \tag{110}$$

where $A(\varepsilon)$ is defined by (96). The same argument as used in proving Theorem 1 leads to

$$\lim_{\varepsilon\to 0} A(\varepsilon)=\mathbf{H}_1(\xi). \tag{111}$$

Thus to prove Theorem 5 it suffices to show that

$$\lim_{\varepsilon\to 0}\Delta(\varepsilon)=0. \tag{112}$$

Let $\delta>0$ be arbitrary and let $n_1(\delta)$ denote the number of those $d_k(\varepsilon)$ for which $d_k(\varepsilon)<\frac{\varepsilon}{1+\delta}$. Then we have

$$1=\sum_{k=1}^{n(\varepsilon)} d_k(\varepsilon)\leqq n_1(\delta)\frac{\varepsilon}{1+\delta}+(n(\varepsilon)-n_1(\delta))\varepsilon$$

which implies

$$n_1(\delta)\leqq\frac{(n(\varepsilon)\varepsilon-1)(1+\delta)}{\varepsilon\delta}.$$

It follows that

$$\sum_{d_k(\varepsilon)<\frac{\varepsilon}{1+\delta}} d_k(\varepsilon)\leqq\frac{n_1(\delta)\varepsilon}{1+\delta}\leqq\frac{n(\varepsilon)\varepsilon-1}{\delta}.$$

Thus, if $E(\varepsilon,\delta)$ denotes the union of those intervals $[x_{k-1}(\varepsilon), x_k(\varepsilon))$ for which $d_k(\varepsilon)=x_k(\varepsilon)-x_{k-1}(\varepsilon)<\frac{\varepsilon}{1+\delta}$, we have by our supposition (105)

$$\lim_{\varepsilon\to 0}\mu(E(\varepsilon,\delta))=0$$

where $\mu(E)$ denotes the Lebesgue measure of the set E. As evidently

$$\Delta(\varepsilon)\leqq\log(1+\delta)+\log C\int_{E(\varepsilon,\delta)} f(x)\,d(x),$$

it follows by the absolute continuity of the measure $\nu(E)=\int_E f(x)\,d(x)$ that

$$0\leqq\overline{\lim_{\varepsilon\to 0}}\,\Delta(\varepsilon)\leqq\log(1+\delta).$$

As $\delta > 0$ is arbitrary, it follows that (112) holds, which proves Theorem 6. We hope to return to the discussion of the ε-entropy with respect to a given subdivision of a general completely bounded metric space in another paper.

(Received 23 February 1959)

References

[1] J. Balatoni and A. Rényi, Remarks on entropy (in Hungarian with English and Russian summaries), *Publications of the Mathematical Institute of the Hungarian Academy of Sciences,* **1** (1956), pp. 9—40. (A german (somewhat abridged) translation of this paper has been printed under the title "Über den Begriff der Entropie" in the collection of papers: *Arbeiten zur Informationstheorie. I* (Math. Forschungsberichte. IV, Deutscher Verlag der Wissenschaften, Berlin, 1957, pp. 117—134).)

[2] A. N. Kolmogorov, On the Shannon theory of information transmission in the case of continuous signals, *IRE Transactions of Information Theory,* **1** (1956), pp. 102—108.

[3] А. Н. Колмогоров, Теория передачи информации (Москва, 1956); german translation under the title "Theorie der Nachrichtenübermittlung", in the collection of papers *Arbeiten zur Informationstheorie. I* (Math. Forschungsberichte. IV, Deutscher Verlag der Wissenschaften, Berlin, 1957, pp. 91—116).

[4] А. Н. Колмогоров, О некоторых асимптотических характеристиках вполне ограниченных метрических пространств, ДАН СССР, **108** (1956), pp. 385—388.