DIMENSION, ENTROPY AND INFORMATION

ALFRÉD RÉNYI

BUDAPEST

§ 1. DISCRETE PROBABILITY DISTRIBUTIONS

Let ξ be a random variable having a distribution of the discrete type $P(\xi = x_k) = P_k$ $(k = 1, 2, \ldots, n)$. Let us denote the distribution of ξ by P, i. e. put $P = \{p_1, p_2, \ldots, p_n\}$. The simplest postulational characterization of the entropy $H(p_1, \ldots, p_n) = \sum_{k=1}^{n} p_k \log \frac{1}{p_k}$ of the distribution P of ξ is due to D. K. FADDEEV [1] (the postulates of [1] are adopted also in [2]). The postulates of FADDEEV are the following:

I. $H(p_1, \ldots, p_n)$ is for each n a symmetric function of its variables.

II. $H(p, 1 - p)$ is a continuous function of p $(0 \leqq p \leqq 1)$.

III. $H(p_1, \ldots, p_n) = H(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$.

IV. $H(\frac{1}{2}, \frac{1}{2}) = 1$.

Faddeev has proved that if $H(p_1, \ldots, p_n)$ satisfies these four postulates, then $H(p_1, \ldots, p_n)$ is the entropy of the distribution $P = \{p_1, \ldots, p_n\}$, i. e. the formula of Shannon

$$H(p_1, \ldots, p_n) = \sum_{k=1}^{n} p_k \log \frac{1}{p_k} \tag{1}$$

holds. (Here and in what follows log denotes the logarithm with respect to the base 2.)

The proof of this assertion consists of the following four steps: A) First it is shown, that putting

$$F(n) = H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \tag{2}$$

we have

$$F(nm) = F(n) + F(m) \tag{3}$$

for every natural n and m i. e. $F(n)$ is a completely additive number-theoretical function.

B) After this it is shown that

$$\lim_{n\to+\infty} (F(n+1) - F(n)) = 0\,. \tag{4}$$

C) The only difficult part of the proof is the third step: the proof that (3), (4) and $F(2) = 1$ (which is nothing else than IV) imply that

$$F(n) = \log n\,. \tag{5}$$

D) From (5) it is easy to deduce (1).

The fact that (3) and (4) together with IV imply (5) has been known previously and is due to P. Erdős [3]. In fact, Erdős proved the following

Theorem 1 (Erdős). *Let $F(n)$ be an additive number-theoretical function, i. e. suppose*

$$F(nm) = F(n) + F(m) \quad \textit{if} \quad (n, m) = 1 \tag{6}$$

where (n, m) denotes the greatest common divisor of n and m. If (4) is fulfilled, then we have

$$F(n) = c \log n \tag{7}$$

with some real constant c.

The proof of Theorem 1 given by Erdős is essentially the same as the proof of step C) in [1]. Recently I have found [4] a new and simple proof of Theorem 1. I reproduce it here because it can be given in a few lines. It runs as follows:

Let p be an arbitrary prime, $\alpha \geqq 1$ and put $k = p^\alpha$. Put

$$G(n) = F(n) - \frac{F(k)\log n}{\log k}\,. \tag{8}$$

Clearly $G(n)$ is also additive and putting

$$\delta_n = G(n+1) - G(n) \tag{9}$$

we have

$$\lim_{n\to+\infty} \delta_n = 0 \tag{10}$$

further

$$G(k) = 0\,. \tag{11}$$

Now let for any positive integer n define n' by

$$n' = \begin{cases} \left[\dfrac{n}{k}\right] & \text{if } p \text{ is not a divisor of } \left[\dfrac{n}{k}\right] \text{ or if } \left[\dfrac{n}{k}\right] = 0\,, \\[2ex] \left[\dfrac{n}{k}\right] - 1 & \text{if } p \text{ divides } \left[\dfrac{n}{k}\right] \neq 0\,. \end{cases}$$

where $[x]$ denotes the integral part of x. Evidently

$$n = kn' + r \quad \text{with} \quad (k, n') = 1 \quad \text{and} \quad 0 \leqq r < 2k\,. \tag{12}$$

It follows that

$$G(n) = G(n') + \sum_{l=kn'}^{n-1} \delta_l\,. \tag{13}$$

Repeating the same for n' instead of n, then for $n'' = (n')'$ etc., we obtain, putting $n^{(0)} = n$, $n^{(1)} = n'$ and $n^{(j+1)} = (n^{(j)})'$ $(j = 1, 2, \ldots)$ that

$$G(n) = G(n^{(r)}) + \sum_{j=1}^{r} \sum_{l=kn^{(j)}}^{n^{(j-1)}-1} \delta_l\,.$$

According to (12) we have $n' \leqq \frac{n}{k}$ and thus $n^{(r)} \leqq \frac{n}{k^r}$; thus we obtain for some $r \leqq \left[\frac{\log n}{\log k}\right] + 1$ that $n^{(r)} = 0$, i. e.

$$G(n) = \sum_{j=1}^{S_n} \delta_{h_j} \tag{14}$$

where $h_1 < h_2 < \ldots < h s_n$ and $S_n \leqq 2k \left(\frac{\log n}{\log k} + 1\right)$. It follows by (10) that

$$\lim_{n \to +\infty} \frac{G(n)}{\log n} = 0 \tag{15}$$

and thus by (8)

$$\lim_{n \to +\infty} \frac{F(n)}{\log n} = \frac{F(k)}{\log k} = \frac{F(p^\alpha)}{\log p^\alpha}\,. \tag{16}$$

Denoting the value of the limit on the left of (16) by c it follows from (16) that

$$F(p^\alpha) = c \log p^\alpha \tag{17}$$

and therefore by (6) we obtain that (7) holds.

Thus Theorem 1. is proved. This enables to simplify considerably the proof of Faddeev's theorem that postulates I (to IV) characterize the entropy completely.

§ 2. ENTROPY AND INFORMATION IN CASE OF DISCRETE CONDITIONAL PROBABILITY DISTRIBUTIONS

Let $S = [\Omega, \mathfrak{A}, \mathfrak{B}, P(A \mid B)]$ be a conditional probability space. (See [5].) If ξ is a discrete random variable on S, and $B \in \mathfrak{B}$, then the conditional distribution of ξ under condition B is an ordinary discrete distribution and thus the conditional entropy $H(\xi \mid B)$ of ξ under condition B can be defined as in the case of ordinary probability spaces. Thus if x_k $(k = 1, 2, \ldots, n)$ are the possible values of ξ and

$$P_k(B) = P \ (\xi = x_k \mid B)$$

then

$$H(\xi \mid B) = \sum_{k=1}^{n} p_k(B) \log \frac{1}{p_k(B)} \, . \tag{18}$$

The question arises whether we may speak about the unconditional entropy of ξ; we want to show here that this is in some cases possible in a certain sense.

Let us consider the following particular conditional probability space Σ: let Ω be the set of all positive integers, $\mathfrak{A}$ the set of all subsets of Ω, $\mathfrak{B}$ the set of all finite and non-empty subsets of Ω and put

$$P(A \mid B) = \frac{\mathfrak{N}(AB)}{\mathfrak{N}(B)} \quad \text{for} \quad A \in \mathfrak{A}, \ B \in \mathfrak{B} \, , \tag{19}$$

where $\mathfrak{N}(C)$ denotes the number of elements of the set C. Let Ω_N denote the set $\{1, 2, \ldots, N\}$ and if ξ is an arbitrary random variable on $\Sigma = [\Omega, \mathfrak{A}, \mathfrak{B}, P(A \mid B)]$ (i. e. $\xi = \xi(n)$ $(n = 1, 2, \ldots)$ an arbitrary number-theoretical function) put

$$H(\xi) = \lim_{N \to +\infty} H(\xi \mid \Omega_N) \tag{20}$$

provided, that the limit on the right of (20) exists. The limit may be called the (unconditional) entropy of ξ.

Let us consider some examples. Let us define $\xi_r = \xi_r(n)$ as the r-th binary digit of the random integer n, i. e. if $n = \sum_{k=0}^{S} \varepsilon_k \, . \, 2^k$ with $\varepsilon_k = 1$ or $\varepsilon_k = 0$ put $\xi_r(n) = \varepsilon_r$ $(r = 0, 1, 2, \ldots)$.

Then clearly

$$\lim_{N \to +\infty} H(\xi_r \mid \Omega_N) = 1 \quad (r = 0, 1, \ldots) \, . \tag{21}$$

Thus each binary digit of a random positive integer carries one unit of information, exactly as each binary digit of a random real number x lying in the interval $[0, 1)$.

We now give an example where the limit (20) does not exist. Let $\xi = \xi(n)$ denote the length of the binary representation of a random integer n, i. e. if $n = \sum_{k=0}^{S} \varepsilon_k \, . \, 2^k$, with $\varepsilon_S = 1$, put $\xi(n) = S + 1$. Then evidently if $n \, . \, 2^{-[\log n]} \to$ $\to 1 + p$ $(0 \leqq p \leqq 1)$ we have

$$\lim_{\substack{n \to +\infty \\ n \, . \, 2^{-[\log n]} \to 1+p}} H(\xi \mid \Omega_n) = p \log \frac{1}{p} + (1 - p) \log \frac{1}{1 - p} \, . \tag{22}$$

Thus the limit (20) does not exist.

If ξ and η are two random variables on the conditional probability space Σ, we may define the information content $I(\xi, \eta/\Omega_N)$ contained in ξ with respect to η (or vice versa) under condition Ω_N, by putting

$$I(\xi, \eta \mid \Omega_N) = H(\xi \mid \Omega_N) + H(\eta \mid \Omega_N) - H(\xi, \eta \mid \Omega_N) . \tag{23}$$

We put further

$$I(\xi, \eta) = \lim_{N \to \infty} I(\xi, \eta \mid \Omega_N) \tag{24}$$

provided that this limit exists and call it the (unconditional) amount of information contained in ξ with respect to η (or vice versa).

For example let $\xi_a = \xi_a(n)$ denote the remainder of the random integer n at the division by a, then (denoting by (a, b) the greatest common divisor of a and b)

$$I(\xi_a, \xi_b) = \lim_{N \to \infty} I(\xi_a, \xi_b \mid \Omega_N) = \log (a, b) .$$

Especially if a and b are relatively prime, we have

$$I(\xi_a, \xi_b) = 0$$

(i. e. ξ_a and ξ_b are independent).

In this example the limit of all three quantities on the right of (23) exist for $N \to + \infty$. There are cases however in which not all these three limits exist, or even all three are tending to $+\infty$, nevertheless $I(\xi, \eta \mid \Omega_N)$ tends to a limit for $N \to + \infty$, i. e. $I(\xi, \eta)$ is defined.

For instance put $\xi = \xi(n) = [\sqrt{n}]$, $\eta = \eta(n) = n - [\sqrt{n}]^2$. Then clearly

$$H(\xi, \eta \mid \Omega_N) = \log N \tag{25}$$

further

$$H(\xi \mid \Omega_N) = \tfrac{1}{2} \log N - 1 + 2 \sum_{k=1}^{[\sqrt{N}]-1} \left(\frac{2k+1}{2\sqrt{N}} \log \frac{2\sqrt{N}}{2k+1} \right) \frac{1}{\sqrt{N}} + O\left(\frac{\log N}{\sqrt{N}} \right) .$$

As the second term is a Riemann-sum of the integral

$$2 \int_0^1 x \log \frac{1}{x} \, \mathrm{d}x = \tfrac{1}{2} \log \mathrm{e} \tag{26}$$

it follows

$$H(\xi \mid \Omega_N) = \tfrac{1}{2} \log N + \tfrac{1}{2} \log \mathrm{e} - 1 + o(1) . \tag{27}$$

On the other hand

$$H(\eta \mid \Omega_N) = \tfrac{1}{2} \log N + 2 \sum_{h=1}^{2[\sqrt{N}]} \left(\frac{h}{2\sqrt{N}} \log \frac{2\sqrt{N}}{h} \right) \frac{1}{2\sqrt{N}} + o(1) .$$

As the second term is again a Riemann-sum of the integral (26) it follows

$$H(\eta \mid \Omega_N) = \tfrac{1}{2} \log N + \tfrac{1}{2} \log \mathrm{e} + o(1) . \tag{28}$$

Thus we obtain from (25), (27) and (28)

$$I(\xi, \eta) = \log e - 1 \,. \tag{29}$$

Therefore if we consider the unique representation of the natural number n in the form

$$n = k^2 + l \quad \text{where} \quad 0 \leqq l \leqq 2k \tag{30}$$

then the knowledge of k gives less than a bit information on l, et vice versa.

The remarks made in this § make it clear that the methods of information theory may be applied with success to number theory. We hope to follow further this line of investigations.

§ 3. ON THE DIMENSION AND ENTROPY OF ABSOLUTELY CONTINUOUS PROBABILITY DISTRIBUTIONS IN EUCLIDEAN SPACES

Let ξ be an arbitrary real valued random variable. Put $\xi_n = [n\xi]$ ($n=1$, $2, \ldots$). If the discrete random variable $[\xi]$ has finite entropy $H([\xi])$ then $H(\xi_n)$ is finite for every n. In this case we define

$$\bar{d}(\xi) = \limsup_{n \to +\infty} \frac{H([n\xi])}{\log n} \tag{31}$$

and

$$\underline{d}(\xi) = \liminf_{n \to +\infty} \frac{H([n\xi])}{\log n} \,. \tag{32}$$

It can be shown [7] that $\bar{d}(\xi) \leqq 1$ and thus

$$0 \leqq \underline{d}(\xi) \leqq \bar{d}(\xi) \leqq 1 \,. \tag{33}$$

If $\underline{d}(\xi) = \bar{d}(\xi)$ we put

$$d(\xi) = \lim_{n \to +\infty} \frac{H([n\xi])}{\log n} \tag{34}$$

and call the number $d(\xi)$ the dimension of the probability distribution of ξ. According to what has been said $0 \leqq d(\xi) \leqq 1$.

If ξ has the dimension d ($0 \leqq d \leqq 1$) further if the limit

$$\lim_{n \to +\infty} (H([n\xi]) - d \log n) = H_d(\xi) \tag{35}$$

exists, we call $H_d(\xi)$ the d-dimensional entropy of ξ (of the distribution of ξ).

In a recent paper [7] I have shown that if $H([\xi])$ exists and if the distribution of ξ is absolutely continuous with density $f(x)$, then it has the dimension **1**,

i. e. $d(\xi) = 1$ and

$$H_1(\xi) = \int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)}\, dx \tag{36}$$

provided that the integral on the right of (36) exists. (Special cases of this theorem have been already proved in [6]).

It is shown in [7] also that the 0-dimensional entropy is identical with the entropy as defined by formula (1); more exactly it has been proved that if the distribution of ξ is of the discrete type, $P(\xi = x_k) = p_k$ $(k = 1, 2, \ldots)$ then ξ has the dimension 0 and if the series $\sum\limits_{k=1}^{\infty} p_k \log \frac{1}{p_k}$ is convergent, we have

$$H_0(\xi) = \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}\,. \tag{37}$$

Thus our definition of entropy is in conformity with the usual one.

It is obvious that the converse of the above assertion holds also, i. e. though there exist of course 0-dimensional not-discrete distributions, but if $d(\xi) = 0$ and $\lim\limits_{n \to +\infty} H([n, \xi])$ exist (and is finite) then ξ has a distribution of the discrete type, and thus (37) holds.

As a matter of fact, if the distribution of ξ is not discrete, it has a continuous component, and thus denoting by $F(x)$ the distribution function of ξ we have

$$F(x) = p\, F_d(x) + (1 - p)\, F_c(x)$$

with $0 < p < 1$ where $F_c(x)$ is a continuous distribution function and $F_d(x)$ a discrete distribution function. It follows as $x \log \frac{1}{x}$ is monotonically increasing for $0 < x < \frac{1}{e}$ that

$$H([n\xi]) \geqq (1 - p)\, \Sigma' \left[F_c\left(\frac{k+1}{n}\right) - F_c\left(\frac{k}{n}\right)\right] \log \frac{1}{F_c\left(\frac{k+1}{n}\right) - F_c\left(\frac{k}{n}\right)}$$

where Σ' indicates that the summation is to be extended for all values of k for which $F\left(\frac{k+1}{n}\right) - F\left(\frac{k}{n}\right) < \frac{1}{e}$, i. e. for all except at most two values of k. Thus if the dimension of ξ is 0 and its 0-dimensional entropy finite, then the same holds for a random variable ξ_c whose distribution function is $F_c(x)$. But this is clearly impossible, because for any $\varepsilon > 0$ there exists an integer $n(\varepsilon)$ such that

$$0 \leqq F_c\left(\frac{k+1}{n}\right) - F_c\left(\frac{k}{n}\right) \leqq \varepsilon \quad \text{for} \quad n \geqq n(\varepsilon)\,.$$

Let k_j $\left(j = 1, 2, \ldots, \left[\frac{1}{\varepsilon}\right] - 1\right)$ denote the uniquely determined least integer for which

$$j\varepsilon \leqq F_c\left(\frac{k_j}{n}\right) < (j+1)\,\varepsilon \quad \text{where} \quad n \geqq n(\varepsilon)\,.$$

Then we have

$$H([n\xi_c]) \geqq \Sigma\left(F_c\left(\frac{k_{2j}}{n}\right) - F_c\left(\frac{k_{2j-2}}{n}\right)\right) \log \frac{1}{F_c\left(\frac{k_{2j}}{n}\right) - F_c\left(\frac{k_{2j-2}}{n}\right)}$$

and thus

$$H([n\xi_c]) \geqq \frac{1}{3} \log \frac{1}{\varepsilon}$$

thus the supposition that $H([n\xi])$ is bounded leads to a contradiction.

Similar results hold for random vectors in euclidean spaces of any dimension. If $\vec{\zeta}$ is an S-dimensional random vector ($S = 2, 3, \ldots$) with the components $\xi_1, \xi_2, \ldots, \xi_S$, we denote by $[n\vec{\zeta}]$ the S-dimensional random vector with the components $[n\xi_1], [n\xi_2], \ldots, [n\xi_S]$. We put as in the case $S = 1$

$$(38) \qquad d(\vec{\zeta}) = \lim_{n \to +\infty} \frac{H([n\vec{\zeta}])}{\log n}$$

provided that this limit exists; we have evidently $0 \leqq d(\vec{\zeta}) \leqq S$. We put further in case $d(\vec{\zeta}) = d$

$$(39) \qquad H_d(\vec{\zeta}) = \lim_{n \to +\infty} \left(H([n\vec{\zeta}]) - d \log n\right)$$

provided that this limit exists. Especially if the distribution of $\vec{\zeta}$ is absolutely continuous with density function $f(\vec{x})$ where $\vec{x} = (x_1, \ldots, x_S)$, we have $d(\vec{\zeta}) = S$ and

$$(40) \qquad H_S(\vec{\zeta}) = \int_{-\infty}^{+\infty} f(\vec{x}) \log \frac{1}{f(\vec{x})}\, \mathrm{d}\vec{x}$$

where $\mathrm{d}\vec{x}$ stands for $\mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_S$.

We restrict ourselves here to prove this result for the special case $s = 1$, under the restrictions that $f(x)$ is bounded and vanishes outside a finite interval, let us say the interval $[0, 1)$. (Concerning the proof for the general case we refer the reader to [7].)

In this case we have

$$(41) \qquad H([n\xi]) - \log n = \sum_{k=0}^{n-1} p_{nk} \log \frac{1}{np_{nk}} = \int_0^1 \varphi_n(x) \log \frac{1}{\varphi_n(x)}\, \mathrm{d}x$$

where $p_{nk} = F\left(\frac{k+1}{n}\right) - F\left(\frac{k}{n}\right)$ ($F(x)$ denotes the distribution function of ξ)

and $\varphi_n(x) = np_{nk}$ for $\frac{k}{n} \leqq x < \frac{k+1}{n}$. Now clearly we have

$$\lim_{n \to +\infty} \varphi_n(x) = f(x) \tag{42}$$

for almost all x. As further $\varphi_n(x) \leqq \sup f(x)$, it follows by the bounded convergence theorem of Lebesque that

$$\lim_{n \to +\infty} (H([n\xi]) - \log n) = \int_0^1 f(x) \log \frac{1}{f(x)}\, dx\,. \tag{43}$$

In the general case the proof is somewhat more complicated, but the idea of the proof is the same.

There exist probability distributions for which $\underline{d}(\xi) + \overline{d}(\xi)$. This may occur only if the distribution of ξ has a singular component. For such distributions there is clearly some connection between the Hausdorff-dimension of the set of points of growth of the distribution function and the information-theoretical dimension of the distribution introduced above. We restrict ourselves to the consideration of the following example. Let us divide the interval $[0, 1)$ into $d_1 \geqq 3$ equal subintervals and attribute to b_1 $(1 < b_1 < d_1)$ intervals I_{j_1} $(j_1 = 1, 2, \ldots, b_1)$ chosen arbitrarily among the d_1 subintervals, the measure $\frac{1}{b_1}$ and to the remaining $d_1 - b_1$ subintervals the measure 0. Let us divide each of the intervals I_{j_1} into $d_2 \geqq 3$ equal subintervals, chose b_2 $(1 < b_2 < d_2)$ among them, denote them by $I_{j_1 j_2}$ $(j_2 = 1, 2, \ldots, b_2)$ and attribute to each the measure $\frac{1}{b_1 b_2}$ and to those not chosen the measure 0. Continue this process with the two sequences b_n and d_n $(1 < b_n < d_n;\ n = 1, 2, \ldots)$. Put $S_n = d_1 d_2 \ldots d_n$ and $T_n = b_1 b_2 \ldots b_n$. Clearly in this way a measure ν is uniquely defined in the interval $[0, 1)$, for which $\nu(I_{j_1 j_2 \ldots j_r}) = \frac{1}{b_1 b_2 \ldots b_r}$. Let ξ be a random variable, such that $P(\xi \in A) = \nu(A)$ for any Borel-subset A of the interval $[0, 1)$. Then clearly

$$\frac{H([S_n \xi])}{\log S_n} = \frac{\log T_n}{\log S_n} = \frac{\sum\limits_{k=1}^{n} \log b_k}{\sum\limits_{k=1}^{n} \log d_k}\,.$$

Thus we have

$$\underline{d}(\xi) \leqq \varliminf_{n \to \infty} \frac{\log T_n}{\log S_n} \leqq \varlimsup_{n \to \infty} \frac{\log T_n}{\log S_n} \leqq \overline{d}(\xi)\,.$$

Moreover it is easy to prove that if $\frac{\log d_{n+1}}{\log S_n} \to 0$

then

$$\underline{d}(\xi) = \varliminf_{n \to \infty} \frac{\log T_n}{\log S_n}\,, \tag{44}$$

and

$$(45) \qquad \overline{d}(\xi) = \underline{\lim}_{n \to +\infty} \frac{\log T_n}{\log S_n} \, .$$

As a matter of fact, if $S_n \leqq N < S_{n+1}$ we have

$$H([S_n \xi]) \leqq H([N\xi]) \leqq H([S_n \xi]) + \log (d_{n+1} + 2) \, .$$

Thus especially (in case $d_n = 3$, $b_n = 2$) we obtain that the dimension of the "uniform distribution on Cantor's ternary set" exists and is equal to $\frac{\log 2}{\log 3}$.*)

By means of formulae (44) and (45) it is easy to construct *distributions whose dimension does not exist*. For this purpose we have only to choose such sequences b_n and d_n that the limit $\lim_{n \to \infty} \frac{\log T_n}{\log S_n}$ should not exist. (Take e. g. $d_n = 5$ $b_n = 3 + (-1)^{[\log n]}$).

Let us consider now the Hausdorff-dimension [8] of the set of those points x which belong to an infinity of intervals $I_{j_1 j_2 \ldots j_r}$. Clearly if $|I|$ denotes the length of the interval I,

$$\Sigma \, |I_{j_1 j_2 \ldots j_n}|^{\alpha} = \frac{T_n}{S_n^{\alpha}} \, .$$

Thus if

$$\underline{\lim}_{n \to \infty} \frac{\log T_n}{\log S_n} = \underline{d}(\xi) = \alpha$$

then we have for any $\varepsilon > 0$ and for an infinity of suitable chosen numbers n_k

$$T_{n_k} < S_{n_k}^{\alpha + \varepsilon/2}$$

and thus

$$\Sigma \, |I_{j_1 j_2 \ldots j_{nk}}|^{\alpha + \varepsilon} = \frac{T_{n_k}}{S_{n_k}^{\alpha + \varepsilon}} \leqq \frac{1}{S_{n_k}^{\varepsilon/2}} \, .$$

Denoting by $D(E)$ the Haussdorff-dimension of the set E, we obtain

$$D(E) \leqq d(\xi_E)$$

where ξ_E is a random variable constructed above, which may be called to be uniformly distributed over the set E. Moreover it is not difficult to prove that in this case

$$(46) \qquad D(E) = d(\xi_E) \, .$$

Especially if E is Cantor's set, we have $D(E) = d(\xi_E) = \frac{\log 2}{\log 3}$.

*) This has been proved first by P. Erdős and later independently by Catherine Weisz (oral communication).

§ 4. INFORMATION EXPRESSED BY ENTROPY

We should like to sketch here the connection between the notion of dimension and entropy as defined in this paper and the notion of information as defined by Kolmogorov, Gelfand and Jaglom [9], [10]. They define the information $I(\xi, \eta)$ contained in one of the random variables ξ, η with respect to the other as

$$I(\xi, \eta) = \sup I(f(\xi), g(\eta)) \tag{47}$$

where $f(x)$ and $g(y)$ run over all Borel-measurable functions taking on only a finite number of different values. Thus especially

$$I(\xi, \eta) \geqq H([n\xi]) + H([n\eta]) - H([n\vec{\zeta}]) \tag{48}$$

where $\vec{\zeta}$ is the two-dimensional random vector with the components ξ, η. It follows evidently, that in case $d(\xi)$ and $d(\eta)$ exist, for the finiteness of $I(\xi, \eta)$ it is necessary that we should have

$$d(\vec{\zeta}) = d(\xi) + d(\eta) . \tag{49}$$

As a matter of fact we have always

$$d(\vec{\zeta}) \leqq d(\xi) + d(\eta)$$

and by (48)

$$d(\vec{\zeta}) < d(\xi) + d(\eta)$$

would imply $I(\xi, \eta) = +\infty$.

On the other hand, it can be shown by considerations similar to that used in [7] that if the joint distribution of ξ and η is absolutely continuous with respect to the direct product of the distributions of ξ and η further $d(\xi)$ and $d(\eta)$ exist, then (49) holds. If further not only (49) is fulfilled but putting

$$d = d(\xi) , \quad d' = d(\eta) , \quad H_d(\xi) , \quad H_{d'}(\eta) \quad \text{and} \quad H_{d+d'}(\vec{\zeta})$$

exist, then $I(\xi, \eta)$ exists also and we have

$$I(\xi, \eta) = H_d(\xi) + H_{d'}(\eta) - H_{d+d'}(\vec{\zeta}) . \tag{50}$$

The details will be published elsewhere.

REFERENCES

[1] D. K. Faddejev: *Zum Begriff der Entropie eines endlichen Wahrscheinlichkeitsschemas.* Uspechi Mat. Nauk 11 (1956), 227—231. German translation. Arbeiten zur Informationstheorie I. Deutscher Verlag der Wiss. Berlin,1957, 85—90.

[2] A. Feinstein: *Foundations of information theory.* McGraw-Hill, New York, 1958.

[3] P. Erdős: *On the distribution function of additive functions.* Annals of Mathematics 47 (1946), 1—20.

[4] A. Rényi: *On a theorem of P. Erdős and its application in information theory.* Mathematica, 1959, in print.

[5] A. Rényi: *On a new axiomatic theory of probability.* Acta Mathematica Academiae Scientiarum Hungaricae 6 (1955), 285—335.

[6] A. Rényi-J. Balatoni: *Zum Begriff der Entropie* (in Hungarian). MTA Matematikai Kutató Intézetének Közleményei 1 (1955), 9—40. German translation: Arbeiten zur Informationstheorie, I. Deutscher Verlag der Wissenschaften, Berlin 1957, 117 bis 134.

[7] A. Rényi: *On the dimension and entropy of probability distributions.* Acta Mathematica Academiae Scientiarum Hungaricae, 10 (1959), 193–215.

[8] F. Hausdorff: *Dimension und äusseres Mass.* Mathematische Annalen, 79 (1918), 157—179.

[9] A. N. Kolmogoroff: *Theorie der Nachrichtenübermittlung,* Moscou 1956. German translation: Arbeiten zur Informationstheorie, I. Deutscher Verlag der Wissenschaften, Berlin 1957, 91—116.

[10] I. M. Gelfand-A. M. Jaglom: *Über die Berechnung der Menge an Information über eine zufällige Funktion die in einer anderen zufälligen Funktion enthalten ist.* Uspechi Mat. Nauk 11 (1957), 3—52. German translation: Arbeiten zur Informationstheorie, II. Deutscher Verlag der Wissenschaften, Berlin 1958, 7—56.

HUNGARIAN ACADEMY OF SCIENCES
INSTITUTE OF MATHEMATICS