THE COMPLEXITY OF FINITE OBJECTS AND THE DEVELOPMENT OF THE CONCEPTS OF INFORMATION AND RANDOMNESS BY MEANS OF THE THEORY OF ALGORITHMS

A.K. Zvonkin and L.A. Levin

In 1964 Kolmogorov introduced the concept of the complexity of a finite object (for instance, the words in a certain alphabet). He defined complexity as the minimum number of binary signs containing all the information about a given object that are sufficient for its recovery (decoding). This definition depends essentially on the method of decoding. However, by means of the general theory of algorithms, Kolmogorov was able to give an invariant (universal) definition of complexity. Related concepts were investigated by Solomonoff (U.S.A.) and Markov. Using the concept of complexity, Kolmogorov gave definitions of the quantity of information in finite objects and of the concept of a random sequence (which was then defined more precisely by Martin-Löf). Afterwards, this circle of questions developed rapidly. In particular, an interesting development took place of the ideas of Markov on the application of the concept of complexity to the study of quantitative questions in the theory of algorithms. The present article is a survey of the fundamental results connected with the brief remarks above.

Contents

Preliminary remarks 83
§1. Complexity 88
§2. Algorithmic problems and the complexity of solution. 94
§3. Effective random processes 99
§4. Random sequences 107
§5. The concept of the quantity of information 115
Index of terms and notation 121
Guide to the literature 122
References. 122

Preliminary remarks

In writing this article, apart from the literature quoted, we have used basically material from lectures of Kolmogorov, from a specialist course by Petri and Kanovich, and also the seminar of Dushski and Levin. We are deeply indebted to Andrei Nikolaevich Kolmogorov, who helped us greatly in editing all the preliminary versions of this article; without his

constant support the paper could not have been written at all. Highly valuable for us was the constant contact and discussion of results with M.I. Kanovich and N.V. Petri, for which we are very grateful. We are also very grateful to A.B. Sosinski, who read the whole manuscript and made many valuable remarks. We would also like to thank V.N. Agafanov, Ya.M. Barzdin', A.N. Kolodie, P. Martin-Löf, L.B. Medvedovski, B.A. Uspenski, J.T. Schwartz and all participants in the seminar of A.A. Markov for valuable discussions.

1. **Some definitions and notation.** We shall investigate *words* in the alphabet $\{0,1\}$, i.e. finite sequences of zeros and ones. We establish a one-to-one correspondence between words and the *natural numbers*:

$$\begin{aligned}
\Lambda &\leftrightarrow 0\\
0 &\leftrightarrow 1\\
1 &\leftrightarrow 2\\
00 &\leftrightarrow 3\\
01 &\leftrightarrow 4\\
10 &\leftrightarrow 5\\
11 &\leftrightarrow 6\\
000 &\leftrightarrow 7\\
001 &\leftrightarrow 8\\
&\ldots\ldots
\end{aligned}$$

(Λ is the empty word), and from now on we shall not distinguish between these objects, using arbitrarily either of the terms "word" or "number". We denote them, as a rule, by small Latin letters, the set of all word-numbers being denoted by S.

If a word y is placed to the right of a word x, we get another word which will be denoted by xy. We also have to be able to write the ordered pair of words (x, y) as one word. In order not to introduce special separating signs (like the comma), we agree that if $x = x_1x_2 \ldots x_n$ ($x_i = 0$ or 1), then

$$\bar{x} = x_1x_1x_2x_2 \ldots x_nx_n01. \tag{0.1}$$

Then from the word $\bar{x}y$ we can unambiguously recover both x and y. We denote by $\pi_1(z)$ and $\pi_2(z)$ the functions for which $\pi_1(\bar{x}y) = x$, $\pi_2(\bar{x}y) = y$; if a word z is not representable in the form $\bar{x}y$, then $\pi_1(z) = \Lambda$, $\pi_2(z) = \Lambda$.[1]

The *length* $l(x)$ of a word x denotes the number of symbols in x; $l(\Lambda) = 0$. Obviously,

$$l(xy) = l(x) + l(y), \tag{0.2}$$

$$l(\bar{x}) = 2l(x) + 2. \tag{0.3}$$

We shall denote by $d(A)$ the *number of elements in the set* A. Evidently,

$$d\{x:\ l(x) = n\} = 2^n, \tag{0.4}$$

$$d\{x:\ l(x) \leqslant n\} = 2^{n+1} - 1. \tag{0.5}$$

---

[1] One could construct a more standard enumeration of the pairs (x, y). However, for us it is important that the property (0.11) holds (see below).

The object of our study is also the space Ω of *infinite binary sequences* (to be denoted by small Greek letters). $\Omega^* = \Omega \cup S$ is the set of all *finite or infinite sequences*. Let $\omega \in \Omega^*$; then the *n-fragment* of ω, denoted by $(\omega)_n$, is defined to be the word consisting of the first n symbols of ω (here, if ω is a word and $l(\omega) \leqslant n$, then by definition $(\omega)_n = \omega$). A sequence $\omega \in \Omega$ is called *characteristic* for a set of natural numbers $A = \{n_1, n_2, \ldots\}$ not containing zero if in this sequence the n_1th, n_2th, ... terms are ones and all the other terms are zeros. The set A for which ω is the characteristic sequence will be denoted by S_ω.

We write Γ_x for the set of all sequences beginning with the word x, that is,

(0.6) $\Gamma_x = \{\omega\colon (\omega)_{l(x)} = x\}$.

These sequences are finite or infinite, or only infinite, depending on whether we are studying Ω^* or Ω, respectively; in each particular case this will be clear from the context. We write $x \subset y$ if $\Gamma_x \supseteq \Gamma_y$ (so that x is a beginning of y). The relation $\subset$ is a partial ordering of S (diagram 1).

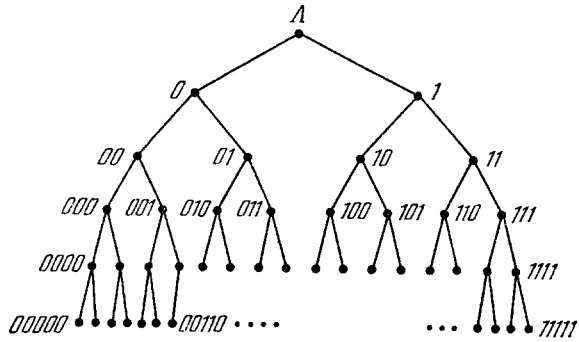


Fig. 1.

Functions defined on the n-fold Cartesian product $S^n = S \times S \times \ldots \times S$ (with the possible exception of standard functions) will be denoted by capital Latin letters, occasionally with an upper index (denoting the number of variables): $F^n = F^n(x_1, \ldots, x_n)$. We always replace the standard phrase: for any admissible values of the digits $y_1, \ldots, y_m$ there exists a constant C such that for all admissible digits $x_1, \ldots, x_n$

$$(0.7)\quad F^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m) \leqslant G^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m) + C,$$

by the shorter phrase (using a new notation):

$$(0.8)\quad F^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m) \preccurlyeq G^{n+m}(x_1, \ldots, x_n; y_1, \ldots, y_m)$$

($y_1, \ldots, y_m$ occur as parameters).

The relation $\succcurlyeq$ is defined analogously; $F \asymp G$ if and only if $F \preccurlyeq G$ and $G \preccurlyeq F$. It is clear that the relations $\preccurlyeq$, $\succcurlyeq$ and $\asymp$ are transitive. Further, it is clear that

$$(0.9)\quad l(x) \succcurlyeq \log_2 x \quad \text{for } x > 0,$$

$$(0.10)\quad l(\bar{x}) \asymp 2l(x),$$

$$(0.11)\quad l(\bar{xy}) \succcurlyeq l(y) \quad (x \text{ occurs as a parameter}),$$

etc.

2. Facts needed from the theory of algorithms. We quote some necessary definitions and theorems from the theory of algorithms. The majority of these facts are proved in any textbook on the theory of algorithms (see, for example, [1] - [4]). The proof of the remaining facts will not present any difficulty to the reader who is familar with one of these textbooks.

Let C^1, O^n, I_m^n be functions defined to take the following values: $C^1(x) = x + 1$, $O^n(x_1, \ldots, x_n) = 0$, $I_m^n(x_1, \ldots, x_n) = x_m$. The $(n+1)$-place function F is said to originate from the n-place function G and from the $(n+2)$-place function H by a *primitive recursion* if for all natural numbers $x_1, \ldots, x_n, y$ we have

$$F(x_1, \ldots, x_n, 0) = G(x_1, \ldots, x_n),$$

$$F(x_1, \ldots, x_n, y+1) = H(x_1, \ldots, x_n, y, F(x_1, \ldots, x_n, y)).$$

We denote by

$$\mu_y(F(x_1, \ldots, x_{n-1}, y) = x_n) \tag{0.12}$$

the least number a for which

$$F(x_1, \ldots, x_{n-1}, a) = x_n. \tag{0.13}$$

Here we agree that the quantity (0.12) is not defined in the following cases:

a) the values $F(x_1, \ldots, x_{n-1}, y)$ are defined for all $y < a$, $y \neq x_n$, but the value $F(x_1, \ldots, x_{n-1}, a)$ is not defined ($a = 0, 1, 2, \ldots$);

b) the values $F(x_1, \ldots, x_{n-1}, y)$ are defined for all $y = 0, 1, 2, \ldots$, $y \neq x_n$.

The value of (0.12) for a given function F depends on the values of $x_1, \ldots, x_{n-1}, x_n$, that is, it is a function of these variables. We say that this function is obtained from F by the *operation of minimization*.

DEFINITION 0.1. A function F is called *partial recursive* if it can be obtained from the functions C^1, O^n, I_m^n by a finite number of operations of *substitution* (that is, superposition), of primitive recursion and of minimization. An everywhere defined partial recursive function is called *general recursive*. A property of numerical n-tuples $\Pi^n(a_1, \ldots, a_n)$ is called a *partial recursive (general recursive) predicate* if there exists a partial recursive (general recursive) function that is equal to zero for all n-tuples satisfying this property, and only for them.

It is easy to verify that the functions $l(x)$, $\pi_1(z)$, $\pi_2(z)$, $F(x) = \overline{x}$, $G(x, y) = xy$ are general recursive.

At the present time, the following scientific hypothesis is generally accepted:

CHURCH'S HYPOTHESIS. *The class of algorithmically computable numerical functions (in the intuitively clear sense) coincides with the class of all partial recursive functions.*

From now on, by quoting the algorithm which computes a certain function, we shall repeatedly assume without proof that it is partial recursive. In fact, because of its bulkiness, we shall not write out the construction required by definition 0.1. The diligent reader, who does not wish to accept Church's hypothesis as true in every case, can always write out such a construction for himself.

REMARK 0.1. It is easy to see that partial recursive functions constructed without the operation of minimization (such functions are called *primitive recursive*) are defined everywhere. Only the operation of minimization can lead to functions that are not defined everywhere. This is because the process of computing the result by minimization (consisting of successive verification of the validity of equation (0.13) for $a = 0, 1, 2, \ldots$) can never stop. We say that the value of the partial

recursive function F^n on the given collection $(x_1, \ldots, x_n)$ is computed *in not more than t steps (operations)* if all operations of minimization involved in constructing F^n were completed on the values of the corresponding parameters a not exceeding t. We often use the concept of the number of steps that were completed by means of the algorithm computing F^n, in the above-mentioned sense.[1]

THEOREM 0.1. *For all partial recursive functions F^n, the following property of the collection $(t; x_1, \ldots, x_n)$ is a general recursive predicate: the value of $F^n(x_1, \ldots, x_n)$ can be computed in not more than t steps.*

DEFINITION 0.2. A partial recursive function $U^{n+1}(i; x_1, \ldots, x_n)$ is called *universal* for all n-place partial recursive functions if for any partial recursive function $F^n(x_1, \ldots, x_n)$ there exists an i such that

$$F^n(x_1, \ldots, x_n) \equiv U^{n+1}(i; x_1, \ldots, x_n). \tag{0.14}$$

The number i is called the *numeral* of F^n with respect to U^{n+1} (a function can have many numerals).

THEOREM 0.2. *For any natural number n there exists a partial recursive function that is universal for all n-place partial recursive functions.*

We define an *enumeration of the set S^n* as any n-tuple of general recursive functions F_i $(i = 1, 2, \ldots, n)$ mapping S onto S^n. A natural number k is called the *numeral* of the n-tuple $(x_1, \ldots, x_n)$ in this enumeration if $F_i(k) = x_i$ for all $i = 1, 2, \ldots, n$. It is evident that the pair of functions $\pi_1(z)$, $\pi_2(z)$ is an enumeration of S^2.

The following definition does not depend on the enumeration.

DEFINITION 0.3. A set $X \subseteq S^n$ is called *enumerable* if the set of numerals of its elements (in the chosen enumeration) is the range of values of some partial recursive function. (Here we say that this function *enumerates X*.)

REMARK 0.2. Any enumerable set can also be enumerated by a general recursive function.

THEOREM 0.3. *Let the predicate Π^{n+k} be partial recursive. Then the set $\{(x_1, \ldots, x_n): \exists a_1, \ldots, a_k \Pi^{n+k}(x_1, \ldots, x_n; a_1, \ldots, a_k)$ is true$\}$ is enumerable.*

The following theorem shows that the family of enumerable sets that depend on the parameters $p_1, \ldots, p_k$ is enumerable *without repetition*.

THEOREM 0.4. *Let $A \subseteq S^{n+k}$ be an enumerable set. Then there exists a partial recursive function $F(t; p_1, \ldots, p_k)$ such that*

a) *for any fixed $p_1, \ldots, p_k$ the set of values of the function $F(t; p_1, \ldots, p_k)$ coincides with the set of numerals of the collection $(x_1, \ldots, x_n)$ such that $(x_1, \ldots, x_n; p_1, \ldots, p_k) \in A$ (the numerals are taken in a certain fixed enumeration of S^n);*

b) *if $t_1 < t_2$ and $F(t_2; p_1, \ldots, p_k)$ is defined, then $F(t_1; p_1, \ldots, p_k)$ is also defined and distinct from $F(t_2; p_1, \ldots, p_k)$.*

---

[1] The number of steps defined in this way is a *signalling* function in the sense of Trakhtenbrot [42].

DEFINITION 0.4. A set $X \subseteq S^n$ is called *solvable* if there exists a general recursive function equal to 0 on X and to 1 on $S^n \setminus X$. The characteristic sequence of a solvable set is called *computable*.

Clearly every solvable set is enumerable.

THEOREM 0.5. *Every infinite enumerable set contains an infinite solvable subset.*

§ 1. Complexity

In this section we introduce the concept of complexity. We derive the simplest evaluations of the quantity of complexity and study the algorithmic properties of this function.

1. Definition. The theorem of optimality. One of the central concepts in this article is the concept of the complexity of a certain text (communication). We define the complexity of a text as the length of the shortest binary word containing all the information that is necessary for recovering the text in question with the help of some fixed method of decoding. More precisely:

DEFINITION 1.1. (Kolmogorov). Let F^1 be an arbitrary partial recursive function. Then the *complexity of the word x with respect to F^1* is:

$$K_{F^1}(x) = \begin{cases} \min l(p): F^1(p) = x, \\ \infty \quad \text{if} \quad \forall p \in S \; F^1(p) \neq x. \end{cases} \tag{1.1}$$

The word p for which $F^1(p) = x$ is called the *code* or *programme* by means of which F^1 recovers the word x.

Such a definition of complexity depends very strongly on the form of F^1. However, the following remarkable theorem permits an invariant definition of this concept. Consequently, the theory as presented in this article could be based on the concept of complexity.

THEOREM 1.1. (Kolmogorov, Solomonoff). *There exists a partial recursive function F^1_0 (called optimal) such that for any other partial recursive function G^1*

$$K_{F^1_0}(x) \preccurlyeq K_{G^1}(x). \tag{1.2}$$

PROOF. See Corollary 1.3.

COROLLARY 1.1. *For any two optimal partial recursive functions F^1 and G^1*

$$K_{F^1}(x) \asymp K_{G^1}(x). \tag{1.3}$$

DEFINITION 1.2. Fix an optimal partial recursive function F^1_0, for example, as in Corollary 1.3 below. Then the *complexity* $K(x)$ of a word x is defined to be $K_{F^1_0}(x)$.

DEFINITION 1.3. (Kolmogorov). The *(conditional) complexity of a word x for a given y with respect to the partial recursive function F^2* is

$$K_{F^2}(x \mid y) = \begin{cases} \min l(p): F^2(p, y) = x, \\ \infty \quad \text{if} \quad \forall p \in S \; F^2(p, y) \neq x. \end{cases} \tag{1.4}$$

THEOREM 1.2. (Kolmogorov, Solomonoff). *There exists a partial recursive function F_0^2 (called optimal) such that for any partial recursive function G^2*

$$K_{F_0^2}(x \mid y) \preccurlyeq K_{G^2}(x \mid y). \tag{1.5}$$

PROOF. Let $U^3(n;\ p,\ y)$ be a partial recursive function that is universal for all two-place partial recursive functions (see, Definition 0.2, Theorem 0.2). We define the function

$$F_0^2(z, y) = U^3(\pi_1(z),\ \pi_2(z),\ y), \tag{1.6}$$

and show that it is optimal. For let G^2 be a partial recursive function, n_{G^2} be any of its numerals (see Definition 0.2), and let

$$K_{G^2}(x \mid y) = l_0, \tag{1.7}$$

so that there exists a programme p_0 for which $G^2(p_0,\ y) = x$, $l(p_0) = l_0$, and the word p_0 has the minimum length of all words p with $G^2(p,\ y) = x$. Then if we substitute $z = \overline{n}_{G^2} p_0$ in (1.6), we get

$$F_0^2(z,\ y) = F_0^2(\overline{n}_{G^2}p_0,\ y) = U^3(\pi_1(\overline{n}_{G^2}p_0),\ \pi_2(\overline{n}_{G^2}p_0),\ y) =$$
$$= U^3(n_{G^2};\ p_0,\ y) = G^2 p_0,\ y) = x,$$

Hence, (1.4), (1.7) and (0.2) imply that

$$K_{F_0^2}(x \mid y) \leqslant l(z) = l(\overline{n}_{G^2}p_0) = l(\overline{n}_{G^2}) + l(p_0) =$$
$$= l_0 + l(n_{G^2}) = K_{G^2}(x \mid y) + l(\overline{n}_{G^2}) \asymp K_{G^2}(x \mid y),$$

since $l(\overline{n}_{G^2})$ does not depend on x and y, but only on G^2.

COROLLARY 1.2. *For any two optimal partial recursive functions F^2 and G^2,*

$$K_{F^2}(x \mid y) \asymp K_{G^2}(x \mid y). \tag{1.8}$$

DEFINITION 1.4. Fix an optimal partial recursive function F_0^2 (for example, as defined by (1.6)). Then the *(conditional) complexity of a word x for a given y* $K(x \mid y)$ is defined to be $K_{F_0^2}(x \mid y)$.

COROLLARY 1.3. *The partial recursive function*

$$F_0^1(p) = F_0^2(p,\ \Lambda) \tag{1.9}$$

is optimal in the sense of Theorem 1.1.

PROOF. We show that $K_{F_0^1}(x) \preccurlyeq K_{G^1}(x)$, where G^1 is an arbitrary partial recursive function. We define $G^2(p,\ y) = G^1(p)$. Then from (1.5) and (1.9) we have $K_{G^1}(x) = K_{G^2}(x \mid \Lambda) \succcurlyeq K_{F_0^2}(x \mid \Lambda) = K_{F_0^1}(x)$, as required.

From now on F_0^1 and F_0^2 will denote optimal functions selected once and for all.

2. Estimates for the quantity of complexity. In this paragraph, we establish the most important estimates for the quantities $K(x)$ and $K(x \mid y)$ that we need in our subsequent investigations.

THEOREM 1.3. (Kolmogorov). *Let A be an enumerable set of pairs $(x,\ a)$, and let $M_a = \{x\colon (x,\ a) \in A\}$. Then*

$$K(x \mid a) \preccurlyeq l(d(M_a)). \tag{1.10}$$

PROOF. Suppose that the partial recursive function $F^2(p, a)$ is computed by the following algorithm we select in the order of enumeration without repetition (see Theorem 0.4) the pth pair of the form (x, a) and take as the value of F^2 the first element of this pair (that is, the word x). It is clear that if $x \in M_a$, then we can find $p \leqslant d(M_a)$ such that $F^2(p, a) = x$; hence, by (1.5) $K(x \mid a) \preccurlyeq K_{F^2}(x \mid a) \leqslant l(d(M_a))$ as required.

REMARK 1.1. For any word y and a finite set M, the number of those $x \in M$ for which

$$K(x \mid y) \leqslant l(d(M)) - m, \tag{1.11}$$

does not exceed 2^{-m+1}. For if $K(x \mid y) \leqslant n$, then a word p can be found of length not exceeding n such that $F^2_0(p, y) = x$. Hence the collection of such words x certainly does not exceed the collection of all programmes p of length at most n; the number of such programmes p is $2^{n+1} - 1$ (see (0.5)). In its turn, $d(M) \geqslant 2^{l(d(M))} - 1$. As a result, the number of words $x \in M$ satisfying (1.11) is at most $\frac{2^{l(d(M))-m+1}-1}{2^{l(d(M))}-1} < 2^{-m+1}$. Thus, the estimate of Theorem 1.3 is exact for the majority of words; this theorem often makes it possible to obtain the best estimates (that is, generally speaking, estimates that cannot be improved) of the complexity of many types of words. We shall use it repeatedly in what follows.

We now prove some properties of absolute (that is, non-conditional) complexity.

THEOREM 1.4. (Kolmogorov). *The following assertions are true:*

(1.12) a) $$K(x) \preccurlyeq l(x)$$

(therefore, $K(x) < \infty$ for all $x \in S$);

b) *the number of words x for which $K(x) < l_0 - m$ and $l(x) = l_0$ does not exceed 2^{-m+1} (so that the estimate (1.12) is exact for the majority of words);*

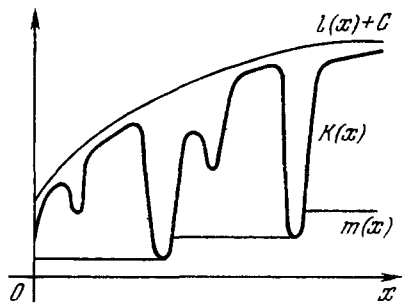
(1.13) c) $$\lim_{x \to \infty} K(x) = \infty$$

(therefore also $\lim\limits_{x \to \infty} m(x) = \infty$), where

(1.14) $$m(x) = \min_{y \geqslant x} K(y),$$

that is, $m(x)$ is the largest monotonic increasing function bounding $K(x)$ from below);

d) *for any partial recursive function $\Phi(x)$ tending monotonically to ∞ from some x_0 onwards, we have $m(x) < \Phi(x)$ (in other words, although $m(x)$ also tends to infinity, it does so more slowly than any partial recursive function that tends to infinity);*



(1.15) e) $$|K(x + h) - K(x)| \preccurlyeq 2l(h)$$

(that is, although $K(x)$ varies all the time between $l(x)$ and $m(x)$, it does so fairly smoothly).

PROOF. (Diagram 2).

a) Let $G^1(x) = x$; then $K_{G^1}(x) = l(x)$ and by Theorem 1.1 $K(x) \preccurlyeq K_{G^1}(x) = l(x)$, as required.

b) This assertion is a trivial corollary of Remark 1.1 (for $y = \Lambda$). We add to this that for any l_0 a word x of length l_0 can be found such that $K(x) \geqslant l_0$ (since the number of texts having length l_0 is 2^{l_0}, and the number of programmes having length less than l_0 is 2^{l_0-1}).

c) By analogy to Remark 1.1, the number of words x such that $K(x) \leqslant a$ does not exceed 2^{a+1}, so that, of course, for any a there exists an x_0 ($x_0 = \max\limits_{K(x)<a} x$) such that $K(x) > a$ for all $x > x_0$, as required.

d) Suppose that the assertion of the theorem is false, so that there exists a partial recursive function $\Phi(x) \leqslant m(x)$ for an infinite set of points x. Then $\Phi(x)$ is defined on an infinite enumerable set U. By Theorem 0.5, U contains an infinite solvable set V. Let us put[1]

$$\Psi(x) = \begin{cases} \Phi(x) \dot{-} 1 & x \in V, \\ \Phi\left(\max\limits_{y \leqslant x,\ y \in V} y\right) \dot{-} 1, & x \notin V. \end{cases}$$

The so constructed function $\Psi(x)$ is general recursive, tends monotonically to infinity, and $\Psi(x) \leqslant m(x)$ on an infinite set of points x. We write $M(a) \max\limits_{K(x) \leqslant a} x$. It is easy to verify that $M(a) + 1 = \min\limits_{m(x)>a} x$. It is not difficult to show that $\max\limits_{\Psi(x) \leqslant a} x \geqslant \min\limits_{m(x)>a} x > M(a)$ on an infinite set of points a, and the function $F(a) = \max\limits_{\Psi(x) \leqslant a} x$ is obviously general recursive. Thus, $F(a) > M(a) = \max\limits_{K(x) \leqslant a} x$ on an infinite set of points a, that is, $K(F(a)) > a$. But by Theorem 1.1, $K(F(a)) \preccurlyeq K_F(F(a)) < l(a)$. Hence there exists a constant C such that $l(a) + C > a$ for an infinitely large set of numbers a, which is impossible.

e) Let p_x be the programme of minimal length for the word x, that is, $F^1_0(p_x) = x$ and $K(x) = l(p_x)$. Then the word $x + h$ can be obtained from the programme $\overline{h}p_x$ by applying to it the function $G^1(z) = F^1_0(\pi_2(z)) + \pi_1(z)$; therefore by (0.2) and (0.10)

$$K_{G^1}(x+h) \leqslant l(\overline{\overline{h}}p_x) = l(\overline{h}) + l(p_x) \asymp 2l(h) + l(p_x) = 2l(h) + K(x).$$

But $K(x + h) \preccurlyeq K_{G^1}(x + h)$, hence $K(x + h) \preccurlyeq K(x) + 2l(h)$, or $K(x + h) - K(x) \preccurlyeq 2l(h)$. Analogously, by applying the function $H^1(z) = F^1_0(\pi_2(z)) \dot{-} \pi_1(z)$ to the word $\overline{h}p_{x+h}$, where p_{x+h} is the programme of the word $x + h$, we obtain

$$K(x) - K(x+h) \preccurlyeq 2l(h).$$

3. Algorithmic properties of complexity. THEOREM 1.5. (Kolmogorov). a) *The function $K(x)$ is not partial recursive, moreover, no partial recursive function $\Phi(x)$ defined on an infinite set of points can coincide with $K(x)$ in the whole of its domain of definition.*

[1] $a \dot{-} b = \max\{a - b; 0\}$; this operation is introduced in order not to go outside the set of natural numbers.

b) *There exists a general recursive function* $H(t, x)$ *monotonically decreasing in* t *such that*

$$\lim_{t\to\infty} H(t, x) = K(x) \tag{1.16}$$

(in other words, although there is no method of computing $K(x)$*, we can nevertheless obtain arbitrarily good upper estimates for this quantity).*

PROOF. a) We select an infinite solvable set V in the domain of definition U of $\Phi(x)$ (see Theorem 0.5). The function $F(m) = \min_{K(x)\geqslant m,\, x\in V} x$ is general recursive (since $K(x) = \Phi(x)$ on V) and takes arbitrarily large values; also $K(F(m)) \geqslant m$ (by construction). On the other hand, $K(F(m)) \leqslant K_F(F(m)) \leqslant l(m)$, hence $m \leqslant l(m)$, which is false.

b) Let C be a sufficiently large constant (such that $K(x) < l(x) + C$). We take the algorithm that computes the function F^1_0 and make it complete in t steps (see Remark 0.1) on all words p of length less than $l(x) + C$. If the word x has not yet been obtained as a result, we put $H(t, x) = l(x) + C$; if it has already been obtained as a result (and possibly not only once), we put $H(t, x)$ equal to the minimum length of the programmes p from which x was obtained. It is clear that $H(t, x)$ is general recursive and monotonically decreasing in t. If we complete more and more steps of the algorithm that computes $F^1_0(p)$ (that is, as $t \to \infty$), we finally obtain x from its "true" programme p_0 of minimum length, that is, we find the complexity of x ($K(x) = l(p_0)$). (True, at no step can we recognise whether this has already happened or not.)

THEOREM 1.6. (Barzdin'). *Let* $f(x)$ *be a general recursive function and* $\lim_{x\to\infty} f(x) = \infty$. *Then the set* $A = \{x: K(x) \leqslant f(x)\}$ *is enumerable (and, in general, the predicate* $\Pi(x, a) \sim [K(x) \leqslant a]$ *is partial recursive). The complement of* A *is infinite, but does not contain any infinite enumerable subset (such sets* A *are called simple).*

PROOF. The assertion $[K(x) \leqslant a]$ is equivalent to $[\exists t: H(t, x) \leqslant a]$ (see Theorem 1.5 b), which proves the first part of the theorem.

Let D be an infinite, enumerable set lying in the complement of A, and suppose that G^1 acts in the following way: it takes the first number $x \in D$, in the order of enumeration without repetition (see, Theorem 0.4) for which $f(x) \geqslant n$ and puts $G^1(n) = x$. It is clear that $K(x) \leqslant K_{G^1}(x) \leqslant l(n)$. But x lies in the complement of A, so that by definition $K(x) \geqslant f(x)$, hence $K(x) \geqslant n$ and $l(n) \geqslant n$, which is false.

4. Majorants of complexity. Obviously, if we know the word x itself and its complexity, then we can effectively (for example, by sorting out) find one of the programmes of least length which code the word x. Moreover, if we know the word x and any number $s \geqslant K(x)$, then we can effectively find one of the programmes of x which, although possibly not the shortest, nevertheless are of length not exceeding s. Since, as follows from Theorem 1.5, we cannot effectively find the complexity, in practice we have to be content with effectively computable (more precisely, partial recursive) functions. These functions are of no less complexity in their whole domain of definition, in other words, they give a value of the code's length which, although not the shortest, yet is effectively computable.

DEFINITION 1.5. We define a *majorant of complexity* as any partial recursive function $\Phi(x)$ for which

$$K(x) \preccurlyeq \Phi(x). \tag{1.17}$$

THEOREM 1.7. (Levin). *A partial recursive function $\Phi(x)$ is a majorant of complexity if and only if*

$$l(d\{x : \Phi(x) = a\}) \preccurlyeq a. \tag{1.18}$$

PROOF. Let Φ be ammajorant of complexity, and let x belong to its domain of definition; $\Phi(x) = a$. By (1.17) a constant C exists such that $K(x) \leqslant \Phi(x) + C$, hence $d\{x: \Phi(x) = a\}$ does not exceed the number of words x such that $K(x) \leqslant a + C$, and so (similarly to Remark 1.1)

$$d\{x : \Phi(x) = a\} \leqslant 2^{a+C+1} \text{ and } l(d\{x : \Phi(x) = a\}) \leqslant a + C + 1,$$

which proves the theorem in one direction.

Now suppose that condition (1.18) holds for a partial recursive function Φ, so that there exists a constant C such that $d\{x: \Phi(x) = a\} \leqslant 2^{a+C}$ for all a. If $\Phi(x) = a$, then the word x can be coded in the following way: let $F(i, a)$ enumerate without repetition all the words y such that $\Phi(y) = a$. (The predicate $[\Phi(x) = a]$ is partial recursive, hence such a function $F(i, a)$ exists; see Theorems 0.4 and 0.3 and Definition 0.1). We write the word i for which $F(i, a) = x$ (it is easy to see that $i \leqslant 2^{a+C}$), and prefix it by the cipher 1, attaching so many zeros on the left that the length of the word becomes $a + C + 1$. From this word it is easy to recover x (to start with, we obtain a by subtracting $C + 1$ from the length of the code; then we find i by throwing away from the left all the zeros and the first 1, thereby giving the word $F(i, a)$). Therefore, $K(x) \preccurlyeq a + C + 1 = \Phi(x) + C + 1$, which proves the theorem in the other direction.

REMARK 1.2. From any partial recursive function $F(x)$ a majorant of complexity can be made by restricting its domain of definition to the set of those x for which $F(x) \geqslant K(x)$. (A priori it is not obvious that the function so obtained is partial recursive; this follows easily, however, from Theorem 1.5b). Hence, in particular, the enumerability of the set of majorants of complexities follows immediately.

In practice, the general recursive majorants of complexity[1] are of special interest, because in the search for a short code of a word it is important to be sure that we shall sooner or later find at least one code. As examples of such general recursive majorants we can take the complexities with respect to any general recursive function (see Definition 1.1).[2] In Theorem 5.1 yet another important example of a majorant of complexity is quoted – the "touched-up" entropy of Shannon.

It is interesting to investigate complexity in so far as it is (up to an additive constant) an exact lower bound of majorants of complexities (see Theorem 1.5b). Hence for a wide class of propositions their statement for complexity is the generalization of their statements for all majorants of complexities. It is remarkable that even in such a strong form these assertions remain true.

---

[1] For details of such functions, see [19] for instance, and also Theorem 2.5 of the present article.

[2] A general recursive function cannot, of course, be optimal.

REMARK 1.3. All results of §1, 3 and 4 and also the definition of majorant of complexity can be transferred without difficulty to the case of conditional complexity $K(x \mid y)$; here, the word y figures as a parameter in all statements and proofs.

§ 2. Algorithmic problems and the complexity of solution

We shall study the behaviour of the complexity of fragments of various infinite binary sequences. With this aim we introduce the concept of the complexity of solution, which is more suitable than $K(x)$ for investigating sequences.

1. Definition and simplest properties. In the preceding section we have developed the apparatus of complexities of those words whose interpretations are complete texts. However, in practice we often have to investigate words representing sequences that are cut short at a more or less arbitrary place, Examples of such words are the approximate value of physical constants, pieces of the text of telegrams, tables of random numbers, cuttings of newspapers up to a given number, etc. It is not interesting to measure the complexity of an algorithm restoring such a word, because even if we possess full information about all sequences, we do not know at what sign the sequence has been cut short. To measure the complexity of a word of known length (that is, assuming an already given truncation place) is not natural either, since it may happen accidentally that the length of the word contains additional information about it. For instance, the binary label of the length could coincide with the beginning of the word. It is far more natural to measure the complexity of the algorithm (or code) which for each number $i \leqslant l(x)$ gives the ith sign of the word in question, in other words, models the activity of the sequence's source up to the ith sign.

DEFINITION 2.1. (Loveland).[1] The *complexity of solution of the word x with respect to the partial recursive function F^2* is defined to be

$$KR_{F^2}(x) = \begin{cases} \min l(p) : \forall i \leqslant l(x)\ F^2(p, i) = x_i, \\ \infty \quad \textit{if no such } p \textit{ exists} \end{cases} \tag{2.1}$$

(here x_i is the ith sign of the word x).

THEOREM 2.1. (Loveland). *There exists an (optimal) partial recursive function G_0^2 such that for any partial recursive function F^2*

$$KR_{G_0^2}(x) \preccurlyeq KR_{F^2}(x). \tag{2.2}$$

The proof is analogous to that of Theorem 1.2.

DEFINITION 2.2. *The complexity of solution $KR(x)$* of the word x is defined as the complexity of its solution with respect to a certain fixed optimal partial recursive function.[2]

The properties of $KR(x)$ are analogous to those of $K(x)$, and the reader will establish them without difficulty. We shall only mention a few of them.

---

[1] Analogous concepts were investigated by Markov (see [15]).

[2] This function will henceforth be denoted by G_0^2.

THEOREM 2.2. (Loveland). a) *If* $x \subset y$, *then*

$$KR(x) \leqslant KR(y). \tag{2.3}$$

b) *An infinite sequence* ω *is computable if and only if the complexity of solution of its fragments is bounded.*

$$K(x) \geqslant KR(x) \geqslant K(x \mid l(x)). \tag{2.4}$$

c)

The proof is obvious.

2. Computable sequences. There is also a less trivial link between the quantities $KR(x)$ and $K(x \mid l(x))$.

THEOREM 2.3. (Kolodii, Levin, Loveland, Mishin). *For* $\omega \in \Omega$, *the quantity* $K((\omega)_n \mid n)$ *is bounded if and only if* $KR((\omega)_n)$ *is bounded.*[1]

PROOF. In one direction the assertion is obvious: if the sequence is computable, then there exists a general recursive function $F^1(n) = (\omega)_n$. We put $F^2(p, n) = F^1(n)$; then $K_{F^2}((\omega)_n \mid n) = l(\Lambda) = 0$, since $F^2(\Lambda, n) = (\omega)_n$; consequently also $K((\omega)_n \mid n) \leqslant 0$, or

$$K((\omega)_n \mid n) \leqslant C. \tag{2.5}$$

Let us prove the converse assertion. Suppose that (2.5) holds. We wish to prove the existence of a procedure which, for each numeral n, would give $(\omega)_n$ as the nth sign of the sequence ω. We write out in a column all words p of length not exceeding C and construct the following table:

	0	1	2	...	n	...
Λ	.	.	.		.	
0	.	$(\omega)_1$	.		.	
1	.	.	.		$(\omega)_n$	
0 0	.	$(\omega)_1$	.		.	
.	.		.		.	
.	.		.		.	
.	.		.		.	
p	. . . . .				$F_0^2(p, n)$	
.	.		.			
.	$(\omega)_0$		$(\omega)_2$			
.						
$\underbrace{11\ldots1}_{C}$						

Corresponding to p, the nth column contains $F_0^2(p, n)$ (see (1.6)) if the function F_0^2 is defined for the pair (p, n). The set of words $F_0^2(p, n)$ appearing in the nth column is denoted by A_n. Each A_n contains not more than 2^{C+1} words, and we always have $(\omega)_n \in A_n$. Let

$$l = \overline{\lim_{n \to \infty}}\, d(A_n).$$

Obviously, the set

$$U = \{n : d(A_n) \geqslant l\}$$

[1] However, as Petri has shown, there is no effective method of evaluating $KR((\omega)_n)$ up to a constant that bounds $K((\omega)_n \mid n)$, so that the former quantity can be very large.

is enumerable and infinite. Here, the definition of l implies that $d(A_n) > l$ for only finitely many numbers n; the largest of these numbers n will be denoted by m_1.

Let the number of sequences ω satisfying (2.5) be k. We denote by m_2 the smallest number such that all m_2-fragments of these k sequences are distinct. In fact, all columns starting with the m_2th must contain at least k words, namely the fragments of these sequences (these fragments are distinct). Hence $k \leqslant l$. Let[1] $m = \max(m_1, m_2)$.

We select from U an infinite solvable subset U' (see Theorem 0.5). Let $V = U' \cap \{n: n > m\}$; obviously, V is also solvable. Renumber the elements of V in increasing order of magnitude: $V = \{n_1, n_2, \ldots\}$. The algorithm solving the ith sequence (in the lexicographic ordering) of our k sequences acts in the following way: suppose that we wish to define the jth sign of the ith sequence. We choose the least $n_r \in V$ such that $n_r > j$, and start filling in the n_rth column (that is, constructing words $F^2_{\mathcal{O}}(p, n_r)$, $l(p) \leqslant C$). As soon as it turns out that l words have already been constructed, we stop: we obtain all words from A_{n_r}. The next step: we choose words of length n_r from A_{n_r}; the set of these words is denoted by B_{n_r}. Next, we construct the set $B_{n_{r+1}}$ similarly, and choose from $B_{n_{r+1}}$ words that are continuations of words from B_{n_r}; the set of these words is denoted by $C_{n_{r+1}}$. Then, from $B_{n_{r+2}}$ we choose words that are continuations of words from $C_{n_{r+1}}$ – they form the set $C_{n_{r+2}}$; $C_{n_{r+3}}$ is the set of words from $B_{n_{r+3}}$ that are continuations of words from $C_{n_{r+2}}$, and so on. We stop when exactly k words occur in the next set C_{n_s}. We are now sure that all words in C_{n_s} are n_s-fragments of sequences satisfying (2.5). We choose from the words in C_{n_s} the ith word in size and find its jth sign. This is what was required.

3. Characteristic sequences of enumerable sets. The complexity of solution of computable sequences is bounded. It is of interest to investigate how the complexity of solution of those sequences increases when they have more complicated algorithmic structure (for example, that of the characteristic sequences of enumerable sets).

THEOREM 2.4. (Barzdin').

a) *For any sequence ω with enumerable S_ω*

$$KR((\omega)_n) \preccurlyeq l(n). \tag{2.6}$$

b) *There exists a sequence with enumerable S_ω such that*

$$KR((\omega)_n) > l(n). \tag{2.7}$$

PROOF. Let $F(x)$ be a function enumerating the set S_ω without repetition (see Theorem 0.4). To restore the word $(\omega)_n$ completely it suffices to give the number s, the last value of the function F (in the order of construction) that does not exceed n. For let $F^2(k, i)$ be obtained in the following way: we compute the values of $F(x)$ until we obtain the

[1] This construction of the algorithm uses the numbers l, k and m. This construction is not effective, because there is no effective procedure for constructing l, k and m (see Footnote on p. 95). We only prove that the required algorithm exists. (An intuitionist might say: "It need not exist".) Therefore, the mere fact of the *existence* of l, k and m is sufficient for us.

number k (if $F(x) \neq k \forall x \in S_\omega$ then $F^2(k, i)$ is not defined). Next we put $F^2(k, i) = 1$ if i has already appeared amongst the values of $F(x)$, and $F^2(k, i) = 0$ otherwise. Then $\omega_i = F^2(s, i)$ for all $i \leqslant n$, hence by (2.1) $KR_{F^2}((\omega)_n) = l(s) \leqslant l(n)$. But by (2.2) $KR((\omega)_n) \preccurlyeq KR_{F^2}((\omega)_n)$. Consequently, (2.6) is true.

b) We put

$$\omega_i = \begin{cases} 1 & \text{if} \quad G_0^2(i, i) = 0, \\ 0 & \text{if} \quad G_0^2(i, i) \neq 0 \text{ or is not defined} \end{cases}$$

(here G_0^2 is as in Theorem 2.1). We claim that for such a sequence (S_ω is obviously enumerable) (2.7) holds. For, suppose that $KR((\omega)_n) \leqslant l(n)$ for some n; then there exists a $p \leqslant n$ such that $G_0^2(p, i) = \omega_i$ for all $i \leqslant n$. In particular, since $p \leqslant n$, it follows that $G_0^2(p, p) = \omega_p$, which contradicts the definition of ω_p.

We quote without proof one result (due to Kanovich) which connects the structure of sequences with enumerable S_ω with their complexity.

The definition of a *process* and related concepts is given on p. . We call a sequence α with enumerable S_α *universal* if for any sequence β with enumerable S_β there exists a rapidly growing (weak tabular) process F such that $\beta = F(\alpha)$. We call a sequence α *sufficiently complicated* if there exists an unbounded general recursive function $F(n)$ such that $KR((\omega)_n) \geqslant F(n)$.

PROPOSITION 2.1. *The concepts of universality and sufficient complexity of a sequence α with enumerable S_α are equivalent.*

COROLLARY 2.1. *Every sufficiently complex sequence α with enumerable S_α is universal with respect to reducibility in the sense of Turing.*

It is remarkable that in the case of sequences with enumerable S_ω the general recursive majorants of complexity (which are really the quantities of practical interest) show a completely different behaviour to complexity itself.[1]

THEOREM 2.5. (Barzdin', Petri). *There exists a sequence ω with enumerable S_ω such that for any general recursive majorant of complexity Φ a constant C can be found such that*

$$\Phi((\omega)_n) \geqslant \frac{n}{C}. \tag{2.8}$$

PROOF. We give a construction of the required sequence. It consists of pieces written one after another having lengths that double at each stage, the length of the ith piece being 2^i. The piece with numeral i is filled out in the following way: consider a partial recursive function F with numeral k (see Definition 0.2), where k is the highest power of 2 dividing i (numerals i having the same k form an arithmetic progression with common difference 2^{k+1}). The ith piece of ω is then the first word x (in the order of recovery by sorting out) of length 2^i for which $F(x) \geqslant l(x) = 2^i$. If there is no such word x (but to check this there is, in general, no algorithm), then let the ith piece consist only of zeros. It is easy to see that S_ω is enumerable.

We say that the ith piece of ω is "defined" by the kth function.

Let $G(x)$ be a general recursive majorant of complexity. Without loss of generality we may suppose that for $G(x)$ strict inequality $\leqslant$ (instead

[1] For further details about this, see [19].

of $\preccurlyeq$) holds in Theorem 1.7.[1] Then that theorem implies that for any i there exists a word x of length 2^i such that $G(yxz) \geqslant l(x) = 2^i$ for all y, z. Consequently, all pieces that are definable by G are non-trivial.

Let us estimate $G((\omega)_n)$. To do this, we investigate the last piece x lying wholly in $(\omega)_n$ that is "definable" by G. The numeral i of this piece satisfies the inequality $i \geqslant l(n) - 2^{k+1} - 1$, where k is the numeral of G (this inequality follows from $2^{i+2^{k+1}} \geqslant \frac{n}{2}$).

Let y and z be words supplementing x to $(\omega)_n$ (so that $yxz = (\omega)_n$; obviously, $l(y) = 2^i - 1$, $l(z) \leqslant 2^{2^i}$). Then $G((\omega)_n) = G(yxz) \geqslant l(x) = 2^i \geqslant$ $\geqslant 2^{l(n)-2^{k+2}} = n/2^{2^{k+2}}$. This proves the theorem if C is chosen to be $2^{2^{k+2}}$; C depends only on G, since k depends only on it (k is the numeral of G).

4. Maximally complex sequences. Solvable and enumerable sets correspond to sets of zero and first rank, respectively, in Kleene's projective classification. Examination of sequences with a more complex set S_ω, for instance of the second rank, that is, expressible by a two-quantifier predicate, shows that there are maximally complex sequences among them. (The complexity of solution of their fragments is asymptotically equal to the length of these fragments.) This fact will be stated more precisely in Theorem 4.5 and Corollary 4.1. There it will be proved that there exists a two-quantifier sequence for which the complexity of its n-fragments differs from n by not more than $4l(n)$. Here we show that we cannot reduce the quantity $4l(n)$ successively. Although for any n there is a word x of length n such that $K(x) > n$ (see the proof of Theorem 1.4b), there is no sequence for which $K((\omega)_n) \geqslant n$. More than that:

THEOREM 2.6 (Martin-Löf). *For any sequence $\omega \in \Omega$ there exist infinitely many numerals n such that*[2]

$$K((\omega)_n) \preccurlyeq n - l(n)$$

PROOF. Among all the words of length n we define a set A_n of "selected" words in the following way (by induction): suppose that we have defined all selected words in the $(n-1)$th row and that the largest of them is y; then we select $2^{n-l(n)}$ words in the nth row beginning with the word following $y1$ (see Diagram 1). If they are not all in this row, then we select the remaining family from the beginning of the next row, and further we begin already to select words from the $(n+2)$th row. It is clear that any sequence has infinitely many selected fragments. (It is easier to see this fact for oneself rather than to explain it to somebody else. It follows from the fact that the number of selected words in the nth row is (as a rule) equal to $2^{-l(n)} \approx 1/n$ and the series $\Sigma 1/n$ diverges.)

---

[1] For this, it is sufficient to increase $G(x)$ by a constant that does not change its asymptotic behaviour.

[2] In fact, Martin-Löf has established a more precise fact, which we quote without proof. Let $F(n)$ be a general recursive function. We say that ω is F-complex if $K((\omega)_n) \geqslant n - F(n)$. Then: a) if $\sum_{n=1}^{\infty} 2^{-F(n)} = \infty$, then F-complex sequences do not exist; b) if $\sum_{n=1}^{\infty} 2^{-F(n)} < \infty$, then two-quantifier F-complex sequences exist, and F-complex sequences form a set of full measure (concerning the measure L, see p. 100).

Let x be a selected word of length n. It is obvious that[1]

$$K(x) \preccurlyeq l\left(d\left\{\bigcup_{k=0}^{n} A_k\right\}\right) \leqslant l\left(\sum_{k=0}^{n} 2^{k-l(k)}\right) \preccurlyeq n - l(n)$$

§3. Effective random processes

This section investigates effective deterministic and non-deterministic processes (algorithms with random entries) producing sequences. The central result is the construction of a universal semi-computable measure and the explanation of its connection with complexity.

1. Definitions. The equivalence of measures. DEFINITION 3.1. An *algorithmic process*, or simply a *process*, is defined to be a partial recursive function F that maps words into words so that if $F(x)$ is defined for the word x and $y \subset x$, then $F(y)$ is also defined and $F(y) \subset F(x)$.

Let ω be an infinite sequence. We apply the process F successively to all fragments of ω as long as this is possible (that is, while F is defined). As a result we obtain fragments of a certain new sequence ρ (possibly finite or even empty),[2] the *result* of applying the process F to ω (so that F maps Ω into Ω^*). In this case the notation $\rho = F(\omega)$ will also be used.

REMARK 3.1. There exists a *universal* process, that is, a partial recursive function $H(i, x)$ such that $H(i, x)$ for any i is a process and that for any process $F(x)$ there exists an i such that

$$H(i, x) \equiv F(x). \tag{3.1}$$

$H(i, x)$ can easily be constructed from a universal partial recursive function $U^2(i, x)$ (see Definition 0.2). Without loss of generality we may assume that

$$H(\Lambda, \Lambda) = \Lambda \tag{3.2}$$

(we shall need this later on). We call two processes F and G *equivalent* if $F(\omega) = G(\omega)$ for any $\omega \in \Omega$.

REMARK 3.2. For any process there exists a primitive recursive process equivalent to it.

DEFINITION 3.2. We say that a process is *applicable* to a sequence ω if the result of its application to ω is an infinite sequence.

REMARK 3.3. Any process on the set of sequences to which it is applicable is a continuous function (with respect to the natural topology of the space of infinite binary sequences).[3]

DEFINITION 3.3. We call a process F *weakly tabular* or *rapidly growing* (*rapidly applicable* to a sequence ω) if there exists a monotone

---

[1] The last inequality follows from the estimate $\sum_{k=0}^{n} 2^{k-l(k)} \leqslant C \cdot 2^{n-l(n)}$.

[2] If $F((\omega)_n)$ for some n is defined and if all $F((\omega)_m)$, $m > n$, coincide with $F((\omega)_n)$ or are not defined, then the *result* will be $F((\omega)_n)$. The empty word is obtained when $F((\omega)_n)$ is not defined or is empty for all n.

[3] In this topology Ω is homoeomorphic to a Cantor perfect set.

unbounded general recursive function $\Phi(n)$ such that for any x (for any x that are fragments of ω) and n for which $l(x) = n$ and $F(x)$ is defined, the length of the word $F(x)$ is not less than $\Phi(n)$. In this case we say that the *speed of growth (of applicability* to ω) of F is not less than $\Phi(n)$.

REMARK 3.4. It is easy to show that a process that is applicable to all $\omega \in \Omega$ is general recursive and rapidly growing. Obviously, the converse is also true.

DEFINITION 3.4. Let P be a probability measure on Ω. We say that a process is *P-regular* if the set of sequences to which it is applicable has P-measure 1.

In order to give an arbitrary measure on the Borel σ-algebra of subsets of Ω it is sufficient to give its values on the sets Γ_x.

DEFINITION 3.5. We call a measure P on Ω *computable* if there exist general recursive functions $F(x, n)$ and $G(x, n)$ such that the rational number

$$\alpha_P(x, n) = \frac{F(x, n)}{G(x, n)} \tag{3.3}$$

approximates the number $P\{\Gamma_x\}$ to within an accuracy of 2^{-n}.

REMARK 3.5. Obviously, if P is computable, then $\alpha_P(x, n+1) + 2^{-(n+1)}$ approximates $P\{\Gamma_x\}$ to within an accuracy of 2^{-n} in excess. Therefore later on, without loss of generality, we shall always suppose that $\alpha_P(x, n)$ is already an approximation in excess, and we shall take $\alpha_P(x, n) - 2^{-n}$ as an approximation falling short of $P\{\Gamma_x\}$ with accuracy 2^{-n}.

We denote by L the *uniform* measure

$$L\{\Gamma_x\} = 2^{-l(x)}. \tag{3.4}$$

This measure corresponds to Bernoulli trials with probability $p = 1/2$. It is also the Lebesgue measure on the interval $[0, 1]$. L is obviously computable.

THEOREM 3.1. (Levin). a) *For any computable measure P and any P-regular process F, the measure*

$$Q\{\Gamma_y\} = P\{\bigcup \Gamma_x : F(x) = y\} \tag{3.5}$$

(that is, the measure according to which the results of F are distributed) is computable.

b) *For any computable measure Q there exists an L-regular process F such that the results of its approximation to sequences distributed according to L are distributed according to Q and such that a process G exists which is the inverse of F (in the domain of definition of $F \circ G$) and is applicable to all sequences except perhaps the solvable ones or those lying in intervals of Q-measure zero.*

PROOF. a) We must be able to compute $Q\{\Gamma_y\}$ with an accuracy of 2^{-n}, or to find[1] $\alpha_Q(y, n)$. We choose m so that

$$P\{\omega : l(F((\omega)_m)) > l(y)\} > 1 - 2^{-(n+1)}$$

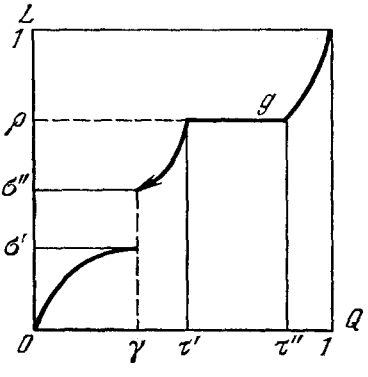
[1] We shall not construct an approximation in excess, but an arbitrary approximation; it is easy to derive an approximation in excess from it (see Remark 3.5).

(such an m exists because F is P-regular, and it is easy to find m effectively). We take all words x of length m such that $y \subset F(x)$ and sum the corresponding measures $P\{\Gamma_x\}$ that are computable to within accuracy $2^{-(m+n+1)}$, in other words, we put

$$\alpha_Q(y, n) = \sum_{x:\, l(x)=m,\ y\subset F(x)} \alpha_P(x,\ m+n+1). \tag{3.6}$$

Then our error $|\alpha_Q(y, n) - Q\{\Gamma_y\}|$ does not exceed $2^{-(n+1)} + 2^m \cdot 2^{-(m+n+1)} = 2^{-n}$ (since there are at most 2^m words x over which the summation was carried out), as required.

b) We regard binary sequences as real numbers in the interval $[0, 1]$ (a sequence is the binary expansion of the number corresponding to it). All cases when this can lead to ambiguity (because the expansion into such sequences of binary rational numbers is non-unique) will be discussed separately. Fig. 3 (where the abscissae and ordinates are distributed according to to Q and L, respectively) shows the distribution function g corresponding to Q. As is well known, if a random variable ξ is uniformly distributed on $[0, 1]$, then $g^{-1}(\xi)$ is distributed according to Q. Our construction will be based on this idea.



I. We construct the process F by inducing Q from L (in fact, this will be the process of calculating g^{-1}; for such a calculation to be possible, it is essential that Q is calculable). Let α be a sequence and $(\alpha)_n$ its n-fragment. With reference to it we find an approximation (with accuracy 2^{-n}) of the number α with deficiency α'_n and excess α''_n. We examine all words y of length n and calculate for each of them the measure $Q\{\Gamma_y\}$ with accuracy 2^{-2n} and excess (that is, $\alpha_Q(y, 2n)$). We select those words z of length n for which

$$\sum_{y\geqslant z} (\alpha_Q(y, 2n) - 2^{-2n}) \geqslant 1 - \alpha''_n \tag{3.7}$$

(the sum on the left is an approximation to $Q\{\bigcup_{y\geqslant z} \Gamma_y\}$ with accuracy 2^{-n} and deficiency) and

$$\sum_{y\leqslant z} \alpha_Q(y, 2n) \geqslant \alpha'_n \tag{3.8}$$

(the sum on the left is an approximation to $Q\{\bigcup_{y\leqslant z} \Gamma_y\}$ with accuracy 2^{-n} and excess). We choose the longest common fragment of all the selected words z and take it as the value of F on $(\alpha)_n$.

II. By (3.7) and (3.8), the sets $\bigcup \Gamma_z$ are intervals containing (for every n) the g-inverse image of the point α. Hence, if the process F is applicable to α, its result will be $g^{-1}(\alpha)$ (we regard γ as the inverse image of points $\alpha \in [\sigma', \sigma'']$, see Diagram 3). To prove that F is the required process we need only show that it is L-regular.

1) Suppose that α lies in an interval of type $[\sigma', \sigma'']$ corresponding to a unique sequence γ of positive measure. If α lies *inside* $[\sigma', \sigma'']$, since the 2^{-n} neighbourhood of $[\alpha_n', \alpha_n'']$ lies completely inside $[\sigma', \sigma'']$, the set of selected words z then consists of the unique word which is an n-fragment of the required sequence γ. Consequently, the process F is applicable to α. In general, F may not be applicable to the end-points of $[\sigma', \sigma'']$.

2) Suppose now that α does not lie in an interval of type $[\sigma', \sigma'']$. Then it follows from (3.7) and (3.8) that $Q\{\bigcup \Gamma_z\} \to 0$ as $n \to \infty$. Hence, if α is not a point of type ρ corresponding to an interval of measure zero, then the intervals $\bigcup \Gamma_z$ themselves shrink to a single point β, namely the g-inverse image of α. Therefore the length of the longest common fragment of the selected words z tends to infinity except, possibly, when β is a binary rational, because if $\beta = m/2^k$, then the intervals $\bigcup \Gamma_z$ can always contain both sequences lying to the left of $m/2^k$ and hence starting with the word $m-1$, and sequences lying to the right of $m/2^k$ and hence starting with the word m. In this case, the longest common fragment of all selected words z is of length less than k).

Thus, F can only be inapplicable to sequences of type ρ, σ' and σ'' (see Diagram 3), and also to sequences having binary rational inverse images. It is clear that the set of such sequences is at most countable. Hence F is L-regular.

III. There is no difficulty in constructing the inverse process: this is the process of calculating the function g. Here G is inapplicable firstly to sequences of type γ having positive measure (such sequences are easily shown to be computable; we do not prove this here, since in Corollary 3.1 a more general result will be proved), and secondly (perhaps) to sequences β on which g takes binary rational values α (in analogy to II (2)). If F is applicable to these binary rational values α, then our sequences β are computable (as F-images of binary rationals). But if F is inapplicable to α, then (see II) our sequences β are either points of type γ (this case has already been investigated) or they form a whole interval $[\tau', \tau'']$ of Q-measure zero, or they themselves are binary rational (consequently computable). The theorem is now completely proved.

2. Semi-computable measures. DEFINITION 3.6. (Levin). A measure is said to be *semi-computable*[1] if the results of applying an arbitrary (not necessarily regular) process to sequences that are distributed according to a certain computable measure are distributed according to it.

REMARK 3.6. A semi-computable measure in concentrated on the space Ω^*, because an irregular process can also yield infinite sequences with positive probability. By Γ_x we understand (throughout this section) the set of all finite or infinite sequences beginning with the word x.

REMARK 3.7. The results of applying any process to sequences that are distributed according to an arbitrary semi-computable measure are also distributed according to a certain semi-computable measure (since the super-position of two processes is a process), and any semi-computable measure can be obtained by a certain process from a uniform measure (see Theorem 3.1b).

[1] The name "semi-computable" is justified by Theorem 3.2.

THEOREM 3.2. (Levin). *A measure P is semi-computable if and only if there exist general recursive functions $F(x, t)$ and $G(x, t)$ such that the function*

$$\beta_P(x, t) = \frac{F(x, t)}{G(x, t)} \tag{3.9}$$

is monotone increasing in t, and

$$\lim_{t\to\infty} \beta_P(x, t) = P\{\Gamma_x\}. \tag{3.10}$$

PROOF. Let P be a semi-computable measure. Then there exists a process F that obtains this measure from a uniform one. We complete it within t steps on all words y of length not exceeding t. Denoting the result by $F_t(y)$ (if it has not yet been obtained, then $F_t(y) = \Lambda$), we put

$$\beta_P(x, t) = L\{\bigcup \Gamma_y : x \subset F_t(y)\}. \tag{3.11}$$

Conversely, suppose that there exists a function $\beta_P(x, t)$ satisfying the conditions of the theorem. We wish to construct a process F that derives P from a uniform measure. The idea of this construction is simple: roughly speaking, we have to decompose the interval $[0, 1]$ into non-intersecting sets of measure $P\{\Gamma_x\}$, and to add the word x when our uniformly distributed sequence gets into the corresponding set. Now we carry out the construction accurately. Obviously, $P\{\Gamma_x\} \geqslant P\{\Gamma_{x0}\} + \{P\{\Gamma_{x1}\}$, Further, without loss of generality we may assume that $\beta_P(x, t) \geqslant \beta_P(x0, t) + \beta_P(x1, t)$ for all t (whenever this inequality is not satisfied, we can decrease $\beta_P(x0, t)$ and $\beta_P(x1, t)$ proportionally to the extent that the inequality becomes valid; by doing this, condition (3.10) is not infringed). It is easy to construct sets in $[0, 1]$ satisfying the following conditions: to each pair (x, t) there corresponds a set, namely the union of finitely many intervals with rational end-points having Lebesgue measure $\beta_P(x, t)$; here, for words $x \neq y$ of the same length the sets corresponding to (x, t_1) and (x, t_2) do not intersect for any t_1 and t_2; if $x \subset y$, then for every t the set corresponding to (x, t) contains that corresponding to (y, t); for $t_1 < t_2$ and every x the set corresponding to (x, t_2) contains that corresponding to (x, t_1).

The process F acts thus: with respect to z it constructs our sets for all pairs (x, t) such that $l(x) \leqslant l(z)$ and $t \leqslant l(z)$ and it produces the word x of largest length such that z belongs to the set corresponding to (x, t) for some t (obviously there is only one such x, because the sets corresponding to various x are disjoint and $x' \subset x''$ if $z' \subset z''$).

3. A universal semi-computable measure. THEOREM 3.3. (Levin). *There exists a universal semi-computable measure R, that is, one satisfying the following condition: for any semi-computable measure Q a constant C can be found such that*

$$C \cdot R\{\Gamma_x\} \geqslant Q\{\Gamma_x\} \tag{3.12}$$

for any x.[1]

[1] In other words, Q is absolutely continuous relative to R, and the Radon-Nikodym derivative is bounded by C.

PROOF. By Remark 3.1 there exists a universal process $H(i, x)$. We put

$$F(z) = H(\pi_1(z), \pi_2(z)). \tag{3.13}$$

It is easy to show that $F(z)$ is a process (see (3.2)). This process, when applied to uniformly distributed sequences, induces the required measure. For suppose that the process $G(y)$ maps a certain set of sequences into Γ_x. Then $F(z)$ maps into Γ_x the same sequences preceded by the word τ, where i is the numeral of G (that is, $H(i, x) = G(x)$ for all x), and possibly also some other sequences. Therefore the measure cannot decrease by more than C times, where we can take $C = 2^{l(\bar{i})}$.

REMARK 3.8. There is no analogous result for computable measures: amongst all computable measures there is no universal one. This fact is one of the reasons for introducing the concept of a semi-computable measure.

The measure R (if we disregard the multiplicative constant) is "larger" than any other measure, and is concentrated on the widest subset of Ω^*. In mathematical statistics the following problem arises: to clarify with respect to what measure a given sequence can be obtained "randomly". If nothing is known in advance about the properties of the sequence, then the only (weakest) assertion we can make regarding it is that it can be obtained randomly with respect to R. Thus, R corresponds to what we intuitively understand by the words "a priori probability". However, the attempt to apply this concept for the foundation of mathematical statistics comes across difficulties connected with the fact that R is not computable.

The following fact is of interest:

a) there exists a constant C such that the probability (with respect to R) of the non-occurrence of the digit 1 after n zeros is not less than $\frac{1}{n} \cdot \frac{1}{C \log_2^2 n}$;

b) for any constant C the portion of those n for which the probability (with respect to R) of the non-occurrence of the digit 1 after n zeros is larger than $\frac{1}{n} C \log_2^2 n$ does not exceed $1/C$ on any sufficiently large interval $[0, N]$.

Therefore, this probability has order[1] approximately $1/n$.

The proof of this assertion follows easily from (3.14) if we take into account that the complexity of solution of a word consisting of n zeros and one 1 does not exceed $\log_2 n$, and for the majority of such words it is almost equal to $\log_2 n$.

We point to an analogy between the construction of complexity and of a universal semi-computable measure. In fact, these quantities turn out to have a numerical connection.

[1] Observe that this assertion is related only to a universal (a priori) probability. For example, if it is known that the sun has been rising for 10,000 years, this still does not mean that the probability that tomorrow it will not rise is equal to approximately $\frac{1}{3,650,000}$. This would be true if our information about the sun were exhausted by the fact stated.

THEOREM 3.4. (Levin).

$$|KR(x)-(-\log_2 R\{\Gamma_x\})| \preccurlyeq 2\log_2 KR(x). \tag{3.14}$$

PROOF. Let $KR(x) = i$, so that there exists a word p with $l(p) = i$ such that $G_0^2(p, n) = x_n$ for every $n \leqslant l(x)$ (here G_0^2 is as in Theorem 2.1). Then it is easy to construct a process that transforms any sequence beginning with the word $\overline{l(p)}p$ into a sequence beginning with the word x: firstly, it must select $\overline{l(p)}$, restore $l(p)$ from $\overline{l(p)}$ and then, knowing $l(p)$, "read" the word p itself; finally, it must start ascribing the corresponding values $G_0^2(p, n)$ for $n = 1, 2, \ldots$. If this process is applied to uniformly distributed sequences, then the induced measure of Γ_x will not be less than $2^{-l(\overline{l(p)}p)}$. Therefore by Theorem 3.3

$$R\{\Gamma_x\} \geqslant C\cdot 2^{-l(\overline{l(p)}p)},$$

hence

$$-\log_2 R\{\Gamma_x\} \preccurlyeq l(\overline{l(p)}p) = l(\overline{l(p)}) + l(p) \asymp l(p) + 2l(l(p)) = \tag{3.15}$$
$$= i + 2l(i) = KR(x) + 2l(KR(x)).$$

Now let $R\{\Gamma_x\} = q$. We write[1] $l(q) = [-\log_2 q]$. We estimate the complexity of solution of the word x; for this purpose we show that any sign of x can be restored with reference to the information given by the triple of words $l(q)$, k and i (or, what is the same thing, by the one word $\overline{l(q)}\,\overline{k}\,i$), where $k = 0$ or 1 and $i \leqslant 2^{l(q)+1}$. Our algorithm acts in the following manner: beginning with the word $l(q)$ it builds up a tree (see Diagram 1) of words y such that $R\{\Gamma_y\} > 2^{-l(q)-1}$ (to do this we have to compute $\beta_R(y, t)$ for all large values of t and y, and to attach y to the tree as soon as $\beta_R(y, t) > 2^{-l(q)-1}$ for some t). The word x belongs to this set. At each stage of the algorithm we select the totality of "maximal" words in the previously constructed part of the tree, that is, words that have as yet no continuation in the previously constructed part of the tree. It is clear that the number of maximal words does not decrease from step to step, remaining less than $2^{l(q)+1}$. In Diagram 4, let A be the point from which the last "collateral branching" from x descends (see Diagram 4), which illustrates the spreading of the tree of words having sufficiently large measure R; a solid line denotes the tree at that instant when the number of branches at first becomes equal to i (at this moment, the branching occurs at A); a dotted line depicts the tree built up as far as all the signs of x have already been solved.) From then on, the word x goes without branching. To solve x, it suffices firstly to give k, which is 0 or 1 according as x goes "to the left" or "to the right" of A, secondly to give some information with reference to which the algorithm could "find" A. As this information we give the number i of maximal words at that moment when both branches at first spread out from A (in the previously constructed part of the tree). This happens precisely when we attach the second branch in the order they are

[1] Here square brackets denote the integral part of a real number.

obtained and the number of maximal words increases by 1 to i. Moreover, $i \leqslant 2^{l(q)+1}$, or $l(i) \leqslant l(q) + 1$. As a result,

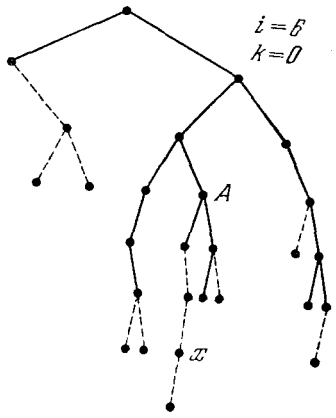


Fig. 4.

$$KR(x) \preccurlyeq l(\overline{l(q)}\,\bar{k}i) \succcurlyeq 2l(l(q)) + l(i) \preccurlyeq 2l(l(q)) + l(q) \succcurlyeq$$
$$\succcurlyeq -\log_2 R\{\Gamma_x\} + 2\log_2(-\log_2 R\{\Gamma_x\}).$$

By (3.15)

$$2\log_2(-\log_2 R\{\Gamma_x\}) \preccurlyeq 2\log_2[KR(x) + 2l(KR(x))] \preccurlyeq 2\log_2 KR(x),$$

hence

$$KR(x) \preccurlyeq -\log_2 R\{\Gamma_x\} + 2\log_2 KR(x); \tag{3.16}$$

(3.15) and (3.16) together give (3.14).

It is interesting to note that by the usual arguments of measure theory it follows that any (not necessarily semi-computable) measure P is almost completely concentrated on the set of those ω for which $\exists\, C$ such that for all n

$$P\{\Gamma_{(\omega)n}\} \geqslant C \cdot R\{\Gamma_{(\omega)n}\}. \tag{3.17}$$

Exactly in the same way, for R-almost all sequences the opposite inequality holds; if P is absolutely continuous relative to R, then the inequality is satisfied for P-almost all sequences. From this it follows that the fact analogous to Theorem 3.4 holds for any semi-computable measure P on the fragments of P-almost any sequence (of course, every sequence having its own constant).

As a corollary to Theorem 3.4 we obtain the well-known theorem of de Leeuw-Moore-Shannon-Shapiro on probabilistic machines.

COROLLARY 3.1. *A sequence ω has positive probability with respect to one (and hence also with respect to a universal) semi-computable measure if and only if ω is computable.*

PROOF. From (3.14) it follows that the measure R of all fragments of ω is larger than a positive number if and only if the complexity of their solution is bounded.

4. Probabilistic machines. The preceding result of Shannon is sometimes interpreted as the impossibility of solving by means of probabilistic machines tasks that are unattainable using deterministic machines. However, the task does not always consist of constructing a certain concrete unambiguously defined object; sometimes the task can have many solutions, and we have to construct only one of them. In such a formulation, obviously, there exist tasks that are unattainable using deterministic machines, but can be solved by means of machines using tables of random numbers (for example, the task of constructing a non-computable sequence).

We say that the task of constructing a sequence having the property Π is *sol-solvable by means of a probabilistic machine* if the universal measure R of such sequences is positive. The following propositions show that such tasks can be solved with arbitrarily large reliability.

PROPOSITION 3.1. (Levin).[1] *Let $A \subseteq \Omega$, $R(A) > 0$. Then for any $\varepsilon > 0$*

[1] We note firstly that the construction of this process with respect to ε is not always effective, and secondly, as N.V. Petri pointed out, that if it is bounded by general recursive processes (and not partial recursive ones) then not every solvable task can be solved by a rapidly growing process.

there exists a weakly tabular (rapidly growing) process with speed of growth $l(F(x)) \geqslant l(x)$, *a process which on application to sequences distributed according to the measure L yields sequences in A with probability at least* $1 - \varepsilon$.

Obviously, for example, one cannot solve the task of obtaining some maximally complex sequence by a more rapidly growing process, because under the application of the process the complexity of words (more precisely, the closely related quantity $[-\log_2 R\{\Gamma_x\}]$) cannot increase. It turns out that when this argument is not essential, the process can be accelerated considerably (that is, the result can be obtained using a smaller number of signs from the tables of random numbers).

PROPOSITION 3.2. (Levin). *Let g be an arbitrary general recursive function. The task of obtaining a sequence from a set A is solvable by means of a process that grows with speed g(n) if and only if there exists a set* $B \subseteq A$, $R(B) > 0$, *such that* $-\log_2 R\{\Gamma_x\} \leqslant n$ *for any sequence* $\omega \in B$, *where* $x = (\omega)_{g(n)}$.

We quote without proof some results concerning the possibility of a solution of standard algorithmic tests by probabilistic machines. The first interesting result of this character is due to Barzdin'. We call an infinite set of natural numbers *immune* if it does not contain any infinite enumerable subset.

PROPOSITION 3.3 (Barzdin'). *There exists an immune set (for example, the complement of the set A in Theorem* 1.6) *such that the task of obtaining a sequence that is characteristic for a certain infinite subset of it is solvable by means of a probabilistic machine.*

The proof of this proposition can easily be obtained from Theorem 1.6 and Corollary 4.1.

An interesting variety of immune sets consists of those sets whose immunity is governed by a too rapid growth of the function that gives, for each i their ith element in order of magnitude; such sets are called *hyperimmune* (more precisely, a set of natural numbers is called hyperimmune if there is no general recursive function F such that $F(i) > x_i$, where x_i is the ith element of the set in order of magnitude).

PROPOSITION 3.4. (Agafonov, Levin). *Whatever the (fixed) hyperimmune set M, the task of obtaining a sequence characteristic for a certain infinite subset of it is not solvable by means of a probabilistic machine.*

However, we have

PROPOSITION 3.5. (Petri). *The task of obtaining a sequence having the property that the set for which it is characteristic is hyperimmune is solvable by means of a probabilistic machine.*

For further details about probabilistic machines, see [23], [25].

§4. Random sequences

1. Definitions. Universal test. The axiomatic construction of probability theory on the basis of measure theory [26] as a purely mathematical discipline is logically irreproachable and does not cast doubts in anybody's mind. However, to be able to apply this theory rigorously in practice its physical interpretation has to be stated clearly. Until recently there was no satisfactory solution of this problem. Indeed, probabiliy is usually interpreted by means of the following arguments: "If we perform many tests, then the ratio of the number of favourable outcomes to the number of tests performed will *always* give a number close to, and in the limit exactly equal to, the probability (or measure) of the event in question. However, to say "always" here would be untrue: strictly speaking, this does not always happen, but only with probability 1 (and for finite series of tests, with probability close to 1). In this way, the concept of the probability of an arbitrary event is defined through the

concept of an event that has probability close to (and in the limit equal to 1), consequently cannot be defined in this manner without an obviously circular argument.

In 1919 von Mises put forward the following way of eliminating these difficulties: according to von Mises there are random and non-random sequences.[1] From the mathematical point of view, random sequences form a set of full measure *and all without exception* satisfy all the laws of probability theory. It is physically possible to assume that as a result of an experiment only random sequences appear.

However, the definition of random sequences proposed by von Mises [27] and later defined more precisely by Wald [28], Church [29] and Kolmogorov [31] turned out to be unsatisfactory. For example, the existence was proved of random sequences, according to von Mises (his so-called *collectives*) that do not satisfy the law of the iterated logarithm [30].

In 1965 Martin-Löf, using ideas of Kolmogorov, succeeded in giving a definition of random sequences free from similar difficulties. Kolmogorov's idea was that one should consider as "non-random" those sequences in which one can observe sufficiently many regularities, where a regularity is defined as *any verifiable property of a sequence inherent only in a narrower class* (of sufficiently small measure). If the "quantity of regularity" is measured according to this traditional logarithmic scale (to base 2) of Shannon, then the last phrase is made more precise in the following way: the measure of the set of sequences containing more than m bits of regularity cannot exceed 2^{-m}.

The choice of scale is not essential for the description of the class of random sequences, and 2^{-m} can be replaced by $1/f(m)$, where $f(m)$ is an arbitrary general recursive monotone unbounded function. However, the choice of scale is a question of the accuracy of measuring the quantity of regularities. But even on a more detailed scale the quantity of regularities could not be measured without obvious arbitrariness, because the theorem on the existence of a universal test (Theorem 4.1) holds only in the logarithmic scale to within an additive constant, and the selection of a less detailed scale would lead to an unjustifiable loss of accuracy.

REMARK 4.1. We stress particularly that by regularities we understand not any rare properties of sequences, but only verifiable ones, that is, we regard as random those sequences which under any algorithmic test and in any algorithmic experiment behave as random sequences.

All the preceding arguments lead us to the following definition.

DEFINITION 4.1. (Martin-Löf). A correct method of proof of P-regularity (where P is a certain probability measure on Ω) or *P-test* is defined to be a function $F(x)$ that satisfies the following conditions:

a) it is general recursive;

b) for $m > 0$

$$P\{\omega: F(\omega) \geqslant m\} \leqslant 2^{-m}, \tag{4.1}$$

where

$$F(\omega) = \sup_n F((\omega)_n). \tag{4.2}$$

[1] We construct the theory in the simplest case, for the space Ω of infinite binary sequences. However, it can easily be generalized (see the small print on pp. 110-111.

The "quantity" of regularities found by a test is taken to be the value of the test. We say that a sequence ω *does not withstand* a P-test F, or that the P-test F *rejects* ω, if $F(\omega) = \infty$.

The meaning of Definition 4.1 a) is conditioned by Remark 4.1. In certain papers tests are investigated for which the condition of computability is replaced by the weaker condition that they can be formulated in a certain theory. That is, these tests also state regularities that cannot be detected yet can somehow be described. Condition (4.1) guarantees that the set of sequences rejected by a P-test has P-measure zero. The converse is also true: for any set of P-measure zero there exists a not necessarily computable function having property (4.1) which rejects all sequences from this set.

Tests can be very varied. However, as in the case of measuring complexity with respect to different partial recursive functions, there is a theorem on the existence of a universal test.

THEOREM 4.1. (Martin-Löf). *For any computable measure P there exists a P-test F (called universal) such that for any P-test G a constant C can be found such that for all $\omega \in \Omega$*

$$G(\omega) \leqslant F(\omega) + C. \tag{4.3}$$

PROOF. We begin by constructing a general recursive function $H^2(i, x)$ such that $H^2(i_0, x)$ is a P-test for any fixed i_0 and that for any P-test G there is an i_0 such that $H^2(i_0, \omega) \geqslant G(\omega) \dot{-} 1$ for all $\omega \in \Omega$. For this purpose we take a universal partial recursive function $U^2(i, x)$ (see Definition 0.2). For each i_0 we transform it in such a way that it becomes a P-test and, if $U^2(i_0, x) + 1$ was already a test, then the suprema over $\omega \in \Omega$ of $U^2(i_0, x)$ are not changed.

We fix i_0. We take all fragments y of the word x and on each fragment we perform $l(x)$ steps of the algorithm computing $U^2(i_0, y)$; we put $G_x(i_0, y)$ equal to the result of applying the algorithm, if the result has already been obtained, and $G_x(i_0, y) = 0$ otherwise. Let $G(i_0, x) = \sup_{y \subset x} G_x(i_0, y)$. Obviously, $G(i_0, x)$ is general recursive and for any $\omega \in \Omega$

$$U^2(i_0, \omega) = G(i_0, \omega). \tag{4.4}$$

However, $G(i_0, x)$ cannot satisfy condition (4.1). To satisfy (4.1) we replace $G(i_0, x)$ by

$$H^2(i_0, x) = \min\{G(i_0, x);\ M(i_0, x)\}, \tag{4.5}$$

where $M(i_0, x)$ is the minimum number m such that for $m + 1$ and $G(i_0, x)$ condition (4.1) is not satisfied "with a reserve" on the accuracy of computation of the measure, that is,

$$M(i_0, x) = \min\Big\{m:\ \sum_{\substack{y: l(y) = l(x) \\ G(i_0, y) \geqslant m+1}} \alpha_P(y, l(x) + m + 2) > 2^{-(m+1)}\Big\}, \tag{4.6}$$

where it can be verified that $m = \max_{y: l(y) = l(x)} G(i_0, y)$ (the sum in (4.5) approximates $P\{\bigcup \Gamma_y:\ l(y) = l(x),\ G(i_0, y) \geqslant m + 1\}$ with surplus and to within $2^{-(m+2)}$). The function $H^2(i_0, x)$ satisfies (4.1) by

construction. Further, if $G(i_0, x) + 1$ satisfies (4.1), then $P\{\bigcup \Gamma_y : l(y) = l(x), G(i_0, y) \geqslant m+1\} \leqslant 2^{-(m+2)}$ (since $G(i_0, y)+1 \geqslant m+2$ for those y) and, consequently, the inequality in (4.6) cannot hold for any m, that is, $G(i_0, x) \equiv H^2(i_0, x)$. Now $H^2(i, x)$ has been constructed.

We shall show that the function

$$F(x) = \max_{i \leqslant l(x)} [H^2(i, x) \dot{-} (i+1)] \tag{4.7}$$

is a universal test. The fact that (4.1) holds for it follows from the inclusion $\{x: l(x) = n, F(x) \geqslant m\} \subseteq \bigcup_{i=1}^{n} \{x: l(x) = n, H^2(i, x) \dot{-} (i+1) \geqslant m\}$ and that $H^2(i, x)$ satisfies (4.1). Finally, if $G(x)$ is a P-test and i_0 is its numeral (see Definition 0.2), then by construction $H^2(i_0, x) \geqslant G(i_0, x) \dot{-} 1$. Hence for words x of length at least i_0 we have $F(x) \geqslant G(i_0, x) \dot{-} (i_0 + 2)$, which implies that for all $\omega \in \Omega$

$$F(\omega) \geqslant G(i_0, \omega) \dot{-} (i_0 + 2). \tag{4.8}$$

Comparing (4.4) and (4.8) we obtain (4.3).

DEFINITION 4.2. (Martin-Löf). We call a sequence ω *random* with respect to a measure P if it withstands any P-test.

With this definition, all random sequences *without exception* satisfy all conceivable effectively verifiable laws of probability theory, since for any such law we can arrange a test that rejects all sequences for which this law does not hold (in other words, the fact that a law is violated is a regularity). A law is understood to mean the assertion that a certain event occurs with probability 1; examples of laws are the strong law of large numbers and the law of the iterated logarithm for sequences of independent trials, the recurrence property of Markov chains, and so on.

REMARK 4.2. According to the theorem just proved, if P is a computable measure, then the randomness of a sequence ω is equivalent to the fact that ω withstands a universal P-test. Thus, for any computable measure the non-randomness of a sequence can be established effectively.

REMARK 4.3. In what follows it will be convenient for us to use a "monotone" universal test, that is, one for which $x \subset y$ implies that $F(x) \leqslant F(y)$. It is easy to obtain it from the constructed test by putting

$$F'(x) = \max_{y \subset x} F(x)\ ^{1}). \tag{4.9}$$

In what follows we shall always assume that the universal test is of this form.

Above we have introduced the concepts of a test of randomness, of a random sequence, of a universal test (and we have proved a theorem on its existence for the case of computable measures) for objects of the simplest type, namely elements of Ω. However, the constructions of Martin-Löf can also be carried out in a more general case. Let T be a topological space with a countable base of open sets x_i ($i = 1, 2, \ldots$), and let P be a measure on the σ-algebra of Borel subsets of T.

---

[1] Here it is obvious that (4.1) is not violated since

$$\sup_n F'((\omega)_n) = \sup_n F((\omega)_n).$$

It will be convenient for us to assume that the elements of the base are numbered in such a way that for any number n of a certain element x of the base we could effectively find a sequence of numbers larger than n that are the numbers of elements of the base whose union is x (for example, we could find another number, larger than n, of x).[1] obviously, such a condition is no restriction of generality, since any enumeration can be altered to an enumeration having this property: replace the (old) number i of each element by the (new) number $(2i + 1).2^k$ $(k = 1, 2, \ldots)$. We say that an element $\omega \in T$ is given if we are given the not necessarily monotone sequence of all numbers i such that $\omega \in x_i$. A general recursive function $F(n)$ is called a *P-test* if

$$P\{\bigcup_{n:\, F(n)\geqslant m} x_n\} \leqslant 2^{-m}. \tag{4.10}$$

We define the *value of the test* F at an element $\omega \in T$ to be

$$F(\omega) = \sup_{n:\, \omega\in x_n} F(n); \tag{4.11}$$

the measure P is called *computable* if for any finite collection of numbers i_k and a number n there is an algorithm computing $P\{\bigcup_k x_{i_k}\}$ with an accuracy of 2^{-n}.

PROPOSITION 4.1. For any computable measure P there exists a universal P-test (*the definition of a universal P-test is the same*).

DEFINITION 4.3. An element $\omega \in T$ is called *P-random* if $F(\omega) < \infty$ for any P-test F.

Obviously, for computable measures this is equivalent to the value of a universal test being finite at ω.

Thus, the concept of a random object has a very general character. Interesting examples of this are the concepts of a random vector, a random element of any function space (of a random process), and so on.

We say that two bases are *equivalent* if for any number i of an element x of one base one can effectively obtain a sequence n_k of numbers of elements of the other base whose union is x.

PROPOSITION 4.2. *The property of an element $\omega \in T$ of being P-random (P is not necessarily computable) is invariant under replacement by an equivalent base.*

REMARK 4.4. If we go over to another enumeration that is non-computably related to the initial one, then we can obtain an inequivalent base. Here the class of random elements can be changed. Example: let $\gamma \in \Omega$ be a random sequence; we renumber the binary words x (they correspond to elements Γ_x of the base) so that the set R of numerals of the fragments of γ becomes solvable, and their length remains a computable function of the numeral. Obviously, the test $F(n)$ that is equal to the length of x if $n \in R$ and equal to zero otherwise, rejects γ. This example shows that the totality of random elements does not depend only on the topologically homogeneous space Ω would be either all random or all non- of a topological homogeneous space Ω would be either all random or all non-random), but also on other structures (for example, connected with the coordinate system.

2. Proper sequences. THEOREM 4.2. (Levin).

a) *For all computable measures P, any P-regular process is applicable to all P-random sequences.*

b) *If P is an arbitrary measure (not necessarily computable), F is a process, Q is the measure induced in F by P, and ω is a P-random sequence to which F is applicable, then its result $F(\omega)$ is Q-random.*

PROOF. a) Let F be a P-regular process that is not applicable to a sequence γ, so that there exists a number (denoted by k) such that the

[1] Omission of this condition would necessitate a more cumbersome definition of of a test.

length of $F(\gamma)$ does not exceed k. This property of our sequence is unique, because the P-measure of such sequences is zero. Hence it is easy to construct a P-test that rejects all sequences ω such that $F(\omega)$ has length at most k. This test acts on the word x as follows: it selects the fragment $(x)_m$ of maximum length such that $l(F((x)_m)) \leqslant k$, then it computes an approximation (with excess to within $2^{-l(x)}$) of the measure of those sequences ω for which $l(F((\omega)_m) \leqslant k$, and it gives as its value for x the integral part of the negative of the logarithm of this measure. Obviously, the value of the test for the words $(\gamma)_n$ tends to infinity. The reader can easily verify that conditions a) and b) of Definition 4.1 are satisfied.

b) Suppose that a Q-test $U(x)$ rejects the sequence $\rho = F(\omega)$, and that G is a general recursive process equivalent to F (see Remark 3.2). Then the the P-test

$$V(x) = U(G(x)) \tag{4.12}$$

(conditions a) and b) of Definition 4.1 are easily verified) rejects ω, i.e. ω is not P-random.

DEFINITION 4.4. We call a sequence *proper* if it is random with respect to a certain computable measure.

All sequences ω with solvable S_ω are proper. It is easy to show that an example of an improper sequence is any sequence ω whose S_ω is the domain of definition of a universal partial recursive function.

THEOREM 4.3. (Levin). *Any proper sequence is either computable or algorithmically equivalent*[1] *to a certain L-random sequence.*

PROOF. Let Q be a computable measure with respect to which our proper sequence ω is random. We show first of all that ω cannot belong to an interval $[\tau', \tau'']$ of Q-measure zero (see Diagram 3). More precisely, there is not a single Q-random sequence in the whole of $[\tau', \tau'']$. For this purpose we construct a Q-test that rejects all sequences from this interval. Let α be a rational number inside the interval. On the word x of length n our test takes as its value the largest number m for which

$$\sum \alpha_Q(y,\ 2n) < 2^{-(m+1)}, \tag{4.13}$$

where the sum is taken over all words y of length n lying between the words x and $(\alpha)_n$ inclusive. Conditions a) and b) of Definition 4.1 are trivially verified, and this test is obviously the required one, because for any sequence β from $[\tau', \tau'']$ the Q-measure of all sequences lying between α and β is zero, and the sum on the left-hand side of (4.13) is the approximation of this measure (with surplus) to within 2^{-n}. That is, the sum tends to zero as $n \to \infty$ and consequently, the value of the test for $(\beta)_n$ tends to infinity as $n \to \infty$.

If ω is not computable, then, since it does not lie in an interval of type $[\tau', \tau'']$, the inverse G of the process F, where F induces Q from L (see Theorem 3.1 b), is applicable to it; we write $G(\omega) = \delta$. The process F is applicable to δ, since it can only be inapplicable to sequences that map into binary rational points (and ω is not a binary rational, because

[1] Two sequences ω' and ω'' are called *algorithmically equivalent* if there exist two processes F and G such that $F(\omega') = \omega''$ and $G(\omega'') = \omega'$.

it is non-computable). F is also applicable to sequences of type ρ which are the image under g of the whole interval (see Theorem 3.1b and Diagram 3), and $\omega \notin [\tau', \tau'']$. Thus, ω and δ are algorithmically equivalent.

It remains to show that δ is L-random. Let U be a universal L-test. Then the Q-test $V(x) = U(G(x))$ (conditions a) and b) of Definition 4.1 are easily verified) rejects all sequences for which the G-results are not L-random; if δ is not L-random, V also rejects ω, that is, ω is not Q-random, which contradicts our assumption.

3. A universal test and complexity. As Theorem 4.3 shows, the study of sequences that are random with respect to an arbitrary computable measure leads to the study of sequences that are random with respect to the uniform measure. We call such sequences simply *random*.

A universal test, if it works on all longer fragments of a sequence, will eventually discover all regularities occurring in the sequence. However, insofar as the universality of the test only appears in the limit, it will only find certain regularities that are concentrated in the initial fragment of the sequence, when it investigates a longer fragment. Then the test takes a small value k on a certain word x, and takes a value $n > k$ on *any* sufficiently long extension of x. It is clear that in this case all these n bits of regularity are defined by x itself and are concentrated in it. We denote by $F(x, n)$ the minimum value of the universal test on words of length n beginning with x. Letting n tend to infinity, we get the quantity of all regularities occurring in x:

$$p(x) = \lim_{n\to\infty} F(x, n)^{1)} \tag{4.14}$$

(according to Remark 4.3, $\lim_{n\to\infty} F(x, n) = \sup_{n\to\infty} F(x, n)$). Obviously (by (4.1)), we always have $p(x) \leqslant l(x)$. The quantity $l(x) - p(x)$, that is, the number of signs in x minus the number of regularities in it, is very similar to complexity as regards its own properties (because of the presence of regularities, these are parasitic signs in the recorded word). It is not computable (since $p(x)$ is not computable), but it can be estimated from above with arbitrary exactness by the function $l(x) - F(x, n)$ (see Remark 4.3 and Theorem 1.5 b). The construction of a universal test is similar to that of an optimal partial recursive function; the portion of words on which $l(x) - p(x)$ takes values significantly less than $l(x)$ is small, etc. It turns out that the quantity $l(x) - p(x)$ and $K(x)$ are also numerically very close.

THEOREM 4.4. (Martin-Löf).

$$|[l(x) - p(x)] - K(x)| \lesssim 4l(l(x)). \tag{4.15}$$

PROOF. Let $l(x) - p(x) \leqslant a$, or $p(x) \geqslant l(x) - a$. This means that on all sufficiently long extensions y of x we have $F(y) \geqslant l(x) - a$, where F is a universal test. From (4.1) it follows that

---

[1] We would remind the reader that we are only considering those regularities which can be demonstrated algorithmically. If $p(x) \leqslant m$ then there exists an infinite sequence $\omega \in \Gamma_x$ in which the universal test discovers at most m regularities. Since the universal test finds all regularities in the limit, there are no other regularities in ω, which means there are none in x, either.

$$L\{\bigcup \Gamma_z: l(z)=l(x),\ p(z)\geqslant l(z)-a\}\leqslant L\{\bigcup \Gamma_y: F(y)\geqslant l(x)-a\}\leqslant 2^{-l(x)+a},$$

and hence the number of words z such that $l(z) = l(x)$ and $p(z) \geqslant l(z) - a$ is at most 2^a, and the collection of such z is enumerable without repetition (see Theorem 0.4). Thus, to find x it is sufficient to give as information the numbers $l(x)$, a and m, where m is the number of x (in order of enumeration) among words z with $l(z) = l(x)$, $p(z) \geqslant l(z) - a$ (so that $m \leqslant 2^a$). This same information can be written in one word: $\overline{l(x)}\,a\,m$. Hence

$$K(x)\leqslant l(\overline{l(x)}\,\bar{a}m)\asymp 2l(l(x))+2l(a)+l(m)\leqslant a+4l(l(x))$$

(because $a \leqslant l(x)$). This inequality holds for all $a \geqslant l(x) - p(x)$, so that $K(x) \leqslant l(x) - p(x) + 4l(l(x))$.

Now we prove this inequality in the opposite direction. For this purpose we construct a test selecting the regularity that consists in the fact that the complexity of a word differs considerably from its length (this is indeed a regularity, because there are few such words – see Theorem 1.4 b). We take the function $H(t, z)$ approximating the complexity from above (see (1.16)). Then the required test is the function

$$G(y)=\max_{i\leqslant l(y)}[i-2-2l(i)+H(l(y),\ (y)_i)]. \tag{4.16}$$

This function takes values at least m only on those sequences ω for which there exists one i such that $K((\omega)_i) \leqslant i - 2 - 2l(i) - m$. By Theorem 1.4 b it follows that the measure of these sequences does not exceed

$$\sum_{i=1}^{\infty}2^{-2l(i)-2-m}\leqslant 2^{-m-1}\sum_{i=1}^{\infty}\frac{1}{i^2}=2^{-m-1}\frac{\pi^2}{6}\leqslant 2^{-m},$$

that is, $G(y)$ satisfies (4.1). The fact that G is general recursive is obvious.

COROLLARY 4.1. *For any random sequence*[1] ω

$$K((\omega)_n)\geqslant n-4l(n) \tag{4.17}$$

COROLLARY 4.2. *A sequence with enumerable* S_ω *cannot be random (with respect to L).*

4. An example of a random sequence. For more complex sets S_ω the situation is different.

THEOREM 4.5. (Martin-Löf). *There exists a random sequence* ω *with respect to L with a set* S_ω *of rank 2 according to Kleene's classification (that is, given by an arithmetical predicate with two quantifiers).*

PROOF. Let A be the set of words having arbitrarily long extensions on which the universal test F takes values not exceeding 1. A is non-empty,

---

[1] As is easy to see from the second part of the proof of Theorem 4.4, $4l(n)$ can be replaced by $2l(n)$ and, in general, by an arbitrary function $F(n)$ such that the series $\sum_{n=1}^{\infty}2^{-F(n)}$ converges computably fast (for example, $F(n) = l(n) + 2l(l(n))$. See also Footnote 2 on p. 98.

since $L\{\omega\colon F(\omega) \geqslant 2\} = 1/4$. Obviously, $\Lambda \in A$, and if $x \in A$, then either $x0$ or $x1$ (or both) belong to A. We define a sequence ω by induction:

$$(\omega)_1 = \begin{cases} 0 & \text{if } 0 \in A, \\ 1 & \text{if } 0 \notin A, \end{cases}$$

$$(\omega)_{n+1} = \begin{cases} (\omega)_n 0 & \text{if } (\omega)_n 0 \in A, \\ (\omega)_n 1 & \text{if } (\omega)_n 0 \notin A \text{ (hence } (\omega)_n 1 \in A). \end{cases}$$

Clearly ω is random, since $F(\omega) \leqslant 1$. We show that ω is of rank 2. To do this we have to construct a solvable predicate $R(n, k, z)$ such that the predicate

$$P(n) \sim \forall k\, \exists z\, R(n, k, z)$$

characterizes S_ω. To construct $R(n, k, z)$ we note that $(\omega)_n$ is the smallest word of length n belonging to A. Therefore the required predicate $R(n, k, z)$ is satisfied by definition if

1) $z = \overline{x}\,\overline{u}\,l$, where x, u, l are words satisfying the following conditions:

2) $l(x) = n$; the last digit of x is 1;

3) $x \subset u$, $F(n) \leqslant 1$, where F is the universal test;

4) $l > n$ and for all pairs of words y, v of lengths n and l, respectively, and such that $y \subset v$ and $y \subset x$, we have $F(v) > 1$.

Theorem 4.5 and Corollary 4.1 make more precise the assertion on the existence of maximally complex sequences of rank 2 that was stated on p. . Of course, the fact that the sequence characterizes a predicate with two quantifiers can be regarded as a regularity. However, it is quite impossible to detect this regularity, and in all algorithmic experiments this sequence is indistinguishable from the remaining mass of random sequences.

§5. The concept of the quantity of information

1. **Definition and simplest properties.** The complexity $K(x)$ intuitively represents the quantity of information necessary for the recovery of a text x. Conditional complexity $K(x \mid y)$ intuitively represents the quantity of information that it is necessary to add to the information contained in the text y, in order to restore the text x. The difference between these quantities is naturally called the quantity of information in y about x.

DEFINITION 5.1. (Kolmogorov). The *quantity of information in y about x* is

$$I(y:x) = K(x) - K(x \mid y). \tag{5.1}$$

REMARK 5.1.

$$I(x:y) \geqslant 0, \tag{5.2}$$

$$|I(x:x) - K(x)| \asymp 0. \tag{5.3}$$

PROOF. We prove (5.2). Let $F^2(p, x) = F^1_0(p)$ (see (1.9)). If $F^1_0(p_0) = y$ and $K(y) = l(p_0)$, then since $F^2(p_0, x) = y$,

$$K(y \mid x) \preccurlyeq K_{F^2}(y \mid x) = K(y).$$

We now prove (5.3). Let $F^2(p, x) = x$. Then also $F^2(\Lambda, x) = x$, hence $K(x \mid x) \preccurlyeq K_{F^2}(x \mid x) = l(\Lambda) = 0$. Noting that $I(x : x) = K(x) - K(x \mid x)$ we obtain the required result.

The following theorem establishes the link between the definitions of quantity of information due to Kolmogorov and to Shannon (more precisely, between the complexity of a word in the sense of Kolmogorov and the entropy of frequency distribution in the sense of Shannon). It turns out that Shannon's entropy is simply the coefficient of the linear part of one of the partial complexities.

THEOREM 5.1. (Kolmogorov). *Let r be a number and suppose that a word x of length $i \cdot r$ consists of i words of length r, where the kth word of length r occurs in x with the frequency q_k ($k = 1, 2, \ldots, 2^r$). Then*

$$K(x) \preccurlyeq i(H(q_k) + \alpha(i)), \tag{5.4}$$

where

$$H(q_k) = -\sum_{k=1}^{2^r} q_k \log_2 q_k \tag{5.5}$$

and

$$\alpha(i) = C_r \frac{\ln i}{i} \to 0 \text{ при } i \to \infty.$$

In the general case, a closer link between entropy and complexity cannot be established. This is indeed natural, since entropy is adapted for studying texts having no regularities other than frequency regularities, that is, for sequences of results independent of the tests. In this special case, we can establish a complete link between the quantities in question (this is done in Theorem 5.3).

PROOF OF THEOREM 5.1. Let x be the mth word in order of magnitude consisting of i words of length r that occur s_k times in it, respectively $\left(q_k = \frac{s_k}{i}, \sum_{k=1}^{2^r} s_k = i\right)$. To find x it is sufficient to give as the information about it the words $m, \underline{s_1}, \ldots, \underline{s_{2^r}}$. All this information can be written in a single word: $p = \underline{s_1}\,\underline{s_2} \ldots \underline{s_{2^r}} m$.

Suppose that the function $F^1(p)$ obtains from this word the word x. Then $K(x) \preccurlyeq 2l(s_1) + \ldots + 2l(s_{2^r}) + l(m)$. We note that m cannot exceed the number of words satisfying the conditions imposed on x, so that $m \leqslant \frac{i!}{s_1! \ldots s_{2^r}!}$. Furthermore, $s_k \leqslant i$. Hence

$$K(x) \preccurlyeq 2^{r+1} l(i) + l\left(\frac{i!}{s_1! \ldots s_{2^r}!}\right). \tag{5.6}$$

Using Stirling's formula $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{\theta_n}{12n}}$, where $|\theta_n| \leqslant 1$, to estimate m we obtain (5.4).

2. The commutativity of information. The classical Shannon quantity of information in one random variable about another satisfies the condition of commutativity, that is, $J(\xi : \eta) = J(\eta : \xi)$. Generally speaking,

there is no exact equation for the Kolmogorov quantity of information in one text about another.

EXAMPLE 5.1. By Remark 1.1 for any l_0 there exists a word x of length l_0 such that $K(x \mid l(x)) \geqslant l(x) - 1$.

By Theorem 1.4b there exist arbitrarily large l_0 such that $K(l_0) \geqslant l(l_0) - 1$. For such a chosen pair of words x and l_0 ($l(x) = l_0$) we have

$$I(x : l_0) = K(l_0) - K(l_0 \mid x) \geqslant l(l_0), \tag{5.7}$$

$$I(l_0 : x) = K(x) - K(x \mid l_0) \leqslant l_0 - l_0 = 0. \tag{5.8}$$

Thus, in certain cases the difference between $I(x : y)$ and $I(y : x)$ can be of order of the logarithm of the complexities of the words in question. However, as Levin and Kolmogorov have shown independently, this order is limited for it and, consequently, if one neglects quantities that are infinitely small in comparison with the information contained in both words, then $I(x : y)$ is commutative all the same.

THEOREM 5.2. (Kolmogorov, Levin).[1]

a) $$|I(x : y) - I(y : x)| \leqslant 12l(K(xy)),$$

b) $$|I(x : y) - [K(x) + K(y) - K(\bar{xy})]| \leqslant 12l(K(xy)).$$

PROOF. a) We only prove the inequality in one direction:

$$I(x : y) \geqslant I(y : x) - 12l(K(xy)). \tag{5.9}$$

The reverse inequality follows from it if by interchanging x and y.

We construct two auxiliary functions. Suppose that the partial recursive function $F^4(n, b, c, x)$ enumerates without repetition words y such that $K(y) \leqslant b$, $K(x \mid y) \leqslant c$. The existence of such a function follows from Theorem 0.4 and Theorem 1.6 (taking into account Remark 1.3). We denote by j the number of such y (j depends non-computably on x, b, c). Then F^4 is defined for all $n \leqslant j$ and only for them. Consequently, the predicate $\Pi(b, c, d, x)$ asserting that the number j defined above exceeds 2^d is obviously equivalent to the assertion that $F^4(2^d, b, c, x)$ is defined, and hence is partial recursive. By analogy with F^4 there exists a function $G^5(m, a, b, c, d)$ that enumerates without repetition all words x such that $K(x) \leqslant a$, $\Pi(b, c, d, x)$. We denote by i the number of these words x (i depends non-computably on a, b, c, d). Obviously, $G^5(m, a, b, c, d)$ is defined for all $m \leqslant i$ and only for them.

Let us proceed with the proof. Let x and y be words with $K(x) = a$, $K(y) = b$, $K(x \mid y) = c$. Then $I(y : x) = a - c$. Further, as was defined above, j is the number of words y' such that $K(y') \leqslant b$ and $K(x \mid y') \leqslant c$ (j depends on x, b, c), and i is the number of words x' such that $K(x') \leqslant a$ and the corresponding number $j' \geqslant 2^{l(j)}$. It is easy to see that $i.2^{l(j)}$ does not exceed the number of pairs (x', y') such that $K(y') \leqslant b$, $K(x' \mid y') \leqslant c$, which in its turn does not exceed 2^{b+c+2}. Hence

$$l(i) + l(j) \leqslant b + c. \tag{5.10}$$

---

[1] By making more accurate estimates, we can improve them slightly. For example, $12l(K(xy))$ can be replaced by $(5 + \varepsilon)l(K(xy))$. It is not known whether the estimate can be reduced to $l(K(xy))$.

Since the word y is given as the value of $F^4(n, b, c, x)$ for some $n \leqslant j$, we see that

$$K(y \mid x) \leqslant l(\bar{b}\bar{c}n) \leqslant 2l(b) + 2l(c) + l(j). \tag{5.11}$$

Further, since the word x is given as the value $G^5(m, a, b, c, d)$ for $d = l(j)$ and some $m \leqslant i$, we have

$$a = K(x) \leqslant l(\bar{a}\bar{b}\bar{c}\bar{d}m) \leqslant 2l(a) + 2l(b) + 2l(c) + 2l(d) + l(i). \tag{5.12}$$

From (5.10) - (5.12) and also from the fact that each of the quantities $l(a)$, $l(b)$, $l(c)$, $l(d) = l(l(j))$ does not exceed $l(K(xy))$, it follows easily that $K(y \mid x) \leqslant b + c - a + 12l(K(xy))$. This implies (5.9).

b) Obviously, $K(x\bar{x}y) \leqslant K(\bar{x}y)$; this implies, by part a) of this theorem, that

$$|I(\bar{x}y : x) - I(x : \bar{x}y)| \leqslant 12l(K(\bar{x}y)),$$

that is,

$$|K(\bar{x}y) - K(\bar{x}y \mid x) - K(x) + K(x \mid \bar{x}y)| \leqslant 12l(K(\bar{x}y)),$$

or

$$|[K(\bar{x}y) - K(x) - K(y)] + K(y) - K(\bar{x}y \mid x) - K(x \mid \bar{x}y)| \leqslant 12l(K(\bar{x}y)),$$

from which we obtain assertion b) of Theorem 5.2 noting that $K(x \mid \bar{x}y) \succ\!\prec 0$ and $|K(\bar{x}y \mid x) - K(y \mid x)| \succ\!\prec 0$.

3. Independent trials. The connection with the probabilistic definition of information. Now we can finally explain the connection between the probabilistic and algorithmic definitions of quantity of information. We recall the former in a form convenient for us (see [39]). If ξ is a random variable taking a finite set of values x_i with probabilities q_i, then we put

$$H(\xi) = -\sum_i q_i \log_2 q_i. \tag{5.13}$$

Let ξ and ψ be two random variables with finite sets of values defined on the same probability space. Then the quantity of information in ξ about ψ is equal to

$$J(\xi : \psi) = H(\xi) + H(\psi) - H(\xi, \psi), \tag{5.14}$$

where (ξ, ψ) is a random vector. If ξ and ψ are random variables with values in Ω (see, however, small print on pp. 110-111), then we put

$$J(\xi : \psi) = \lim_{n \to \infty} J((\xi)_n, (\psi)_n) \tag{5.15}$$

(note that here $\lim\limits_{n \to \infty}$ coincides with $\sup\limits_n$). Suppose that we have two such random variables, jointly distributed according to the measure Q (not necessarily computable). We consider the sequence of independent random vectors $(\xi, \psi)^i$ $(i = 1, 2, \ldots)$, each of which is distributed according to Q. These conditions (independence and identity of distribution according to Q) uniquely define the joint distribution P of the vectors (ξ, ψ). We call a sequence of pairs of infinite binary sequences $(\alpha, \beta)^i$ a sequence of independent trials of the random variables ξ and ψ *random with*

respect to P. We denote by α_n^i and β_n^i the words $\overline{(\alpha^1)_n}\ \overline{(\alpha^2)_n}\ \ldots\ \overline{(\alpha^i)_n}$ and $(\beta^1)_n\ (\beta^2)_n\ \ldots\ (\beta^i)_n$, respectively.

THEOREM 5.3. (Kolmogorov). *If* $(\alpha, \beta)^i$ *is a sequence of independent trials of the random variables* ξ *and* ψ, *then*

$$\lim_{n\to\infty}\lim_{i\to\infty}\frac{I(\alpha_n^i:\beta_n^i)}{i}=J(\xi:\psi). \tag{5.16}$$

PROOF. The assertion of the theorem follows from the equation

$$\lim_{i\to\infty}\frac{I(\alpha_n^i:\beta_n^i)}{i}=J((\xi)_n,(\psi)_n). \tag{5.17}$$

To prove it we note that (5.8) implies

$$\lim_{i\to\infty}\frac{I(\alpha_n^i:\beta_n^i)}{i}=\lim_{i\to\infty}\frac{K(\alpha_n^i)}{i}+\lim_{i\to\infty}\frac{K(\beta_n^i)}{i}-\lim_{i\to\infty}\frac{K(\overline{\alpha_n^i}\beta_n^i)}{i}.$$

By definition,

$$J((\xi)_n:(\psi)_n)=H((\xi)_n)+H((\psi)_n)-H((\xi)_n,(\psi)_n).$$

Hence it is clear that the assertion of the theorem is equivalent to the following.

Let $\theta_1, \theta_2, \ldots$ be a sequence of independent identically distributed random variables taking as values binary words of length r with probability q_k, $k \leqslant 2^r$, and let γ be a binary sequence partitioned into words of length r, which is random with respect to the measure corresponding to the distribution of $\theta_1, \theta_2, \ldots$. Then

$$\lim_{i\to\infty}\frac{K((\gamma)_{i\cdot r})}{i}=H(q_k). \tag{5.18}$$

We prove (5.18). Let x be a word of length $i.r$ consisting of i words of length r occurring in it s_k times, respectively ($\sum_{k=1}^{2^r} s_k = i$). The set of numbers $s_1, \ldots, s_{2^r}$ is called the *set of frequencies of* x. We denote by $h(x)$ the logarithm of the quantity of words having the same set of frequencies as x, that is,

$$h(x)=l\left(\frac{i!}{s_1!\ldots s_{2^r}!}\right).$$

Our sequence γ is random with respect to the measure for independent trials and in each trial the results are obtained with fixed probabilities q_k. Using the strong law of large numbers it is easy to construct, for any $\varepsilon > 0$, $k \leqslant 2^r$, a test rejecting all sequences that have infinitely many fragments in which s_k/i differs from q_k by more than ε. Since γ is random and, consequently, withstands these tests, the limits of s_k/i for its fragments are exactly equal to q_k. From this and Stirling's formula it follows that for the fragments of γ,

$$\lim_{i\to\infty}\frac{h((\gamma)_{i\cdot r})}{i}=H(q_k).$$

We show that

$$\lim_{i\to\infty}\frac{h((\gamma)_{i\cdot r})-K((\gamma)_{i\cdot r})}{i}=0.$$

From Theorem 1.3 we have $K(x\mid \overline{s}_1 \ldots \overline{s}_{2^r}) \preccurlyeq h(x)$, hence $K(x) \preccurlyeq h(x) + 2^{r+1}\cdot l(i)$ (r is fixed); consequently

$$\overline{\lim_{i\to\infty}}\frac{h((\gamma)_{i\cdot r})-K((\gamma)_{i\cdot r})}{i}\geqslant 0.$$

It remains to prove that

$$\overline{\lim_{i\to\infty}}\frac{h((\gamma)_{i\cdot r})-K((\gamma)_{i\cdot r})}{i}\leqslant 0. \tag{5.19}$$

For this purpose we note that since the random variables θ_j are independent and identically distributed, all words with the same set of frequencies $s_1, \ldots s_{2^r}$ are equally probable (their probability being equal to $q_1^{s_1}\ldots q_{2^r}^{s_{2^r}}$). By Remark 1.1 it follows that the quantity of words x with a fixed set of frequencies $s_1, \ldots s_{2^r}$ such that $K(x) \leqslant h(x) - m$ does not exceed 2^{-m}. Consequently, the measure of the set of sequences that start with such words does not exceed 2^{-m}. The measure of the set of sequences that start with a word having the set of frequencies $s_1, \ldots, s_{2^r}$ and satisfying the condition

$$K(x)\leqslant h(x)-2^{r+1}l(i)-m \qquad (i=s_1+\ldots+s_{2^r}), \tag{5.20}$$

does not exceed $2^{-(2^{r+1}l(i)+m)}$. The measure of the set of sequences having *any* fragment satisfying condition (5.20) does not exceed

$$\sum_{(s_1,\ldots,s_{2^r})} 2^{-(2^{r+1}l(s_1+\ldots+s_{2^r})+m)}\leqslant 2^{-m}.$$

Therefore, the test that gives for ω the supremum of the quantity $h(x) - 2^{r+1}l(l(x)) - K(x)$ on all its fragments, satisfies (4.1). It is not difficult to construct its algorithm (this is done as in the second part of the proof of Theorem 4.4). Obviously, this test rejects all sequences that do not satisfy (5.19), and since γ is random, it withstands this test. Hence (5.19) holds as required.

Theorem 5.3 does not only hold for the case of independent trials. Schwartz has raised the question whether a similar fact occurs for arbitrary ergodic stationary processes. A positive answer to this question is given by the following proposition.

PROPOSITION 5.1. (Levin). *Let* $\{\xi_i\}$ ($i = 1, 2, \ldots$) *be any ergodic stationary random process with values* $\xi_i \in \Omega$, P *a measure on its trajectories* $\omega \in \Omega^S$ *given by this process, and* H *its entropy*[1]. *We denote by* $\alpha_n^i(\omega)$ *the word* $(\overline{\xi}_1)_n (\overline{\xi}_2)_n \ldots (\overline{\xi}_i)_n$. *Then for* P*-almost all* ω

$$\lim_{n\to\infty}\lim_{i\to\infty}\frac{K(\alpha_n^i(\omega))}{i}=H.$$

The requirement of ergodicity is not essential here. The only difference is that in the case of a non-ergodic process, the limit under discussion is not a constant H, but a function $f(\omega)$ that is measurable with respect to the σ-algebra of invariant sets of trajectories. It is easy to describe this function. Each invariant

[1] For the definition of entropy of an arbitrary stationary random process, see [40].

set of trajectories A, $P(A) > 0$, can be regarded as an original stationary random process (distributed according to the corresponding conditional probabilities). We denote by $h(A)$ the entropy of this process. It is easy to see that the function $P(A).h(A)$ is additive. Then it has a Radon-Nikodym derivative which is measurable with respect to the σ-algebra of invariant sets. This is the required function $f(\omega)$.

Index of terms and notation

A priori probability 104
Code 88
Complexity 88
–, with respect to F^1 88
–, conditional 88
– –, with respect to F^2 88
–, of solution 94
– –, with respect to G^2 94
Enumeration of S^n 87
Fragment (n-fragment) 85
Function, enumerating a set 87
– – –, without repetition 87
–, general recursive 86
–, optimal 88, 89, 94
–, partial recursive 86
–, primitive recursive 86
–, universal partial recursive 86
Length, of a word 84
Majorant, of complexity 93
Measure, computable 100
–, semi-computable 102
–, uniform 100
–, universal semi-computable 103
Numeral, of a function with respect to U^{n+1} 87
–, of an n-tuple of numbers 87
Predicate, general recursive 86
–, partial recursive 86
–, primitive recursive 86
Programme 88
Process 99
–, applicable to ω 99
–, growth of 100
–, rapidly applicable to ω 100
–, rapidly growing 99
–, regular (P-regular) 100
–, result of applying p. to ω 99
–, speed of applicability of p. to ω 100
–, universal 99
–, weak tabular 99
–es, equivalence of 99
Quantity, of information in one random variable about another 118
– –, in one word about another 115
–, of operations 87
Sequence, characteristic for for a set 85
–, computable 88
–, not withstanding a test 109
–, proper 112
–, random (P-random) 110, 111, 113
–, sufficiently complicated 97
–, universal 97
–s, algorithmic equivalence of 112
Set, enumerable 87
–, hyperimmune 107
–, immune 107
–, simple 92
–, solvable 88
Task, solvable by means of a probabilistic machine 106
Test (P-test) 108, 110
–, rejecting ω 109
–, universal 109
Word 84
$\alpha p(x, n)$ 100
$\beta p(x, t)$ 103
Γ_x 85
$d(A)$ 84
F_0^1 88
F_0^2 89

G_0^2	94	$KR(x)$	94	U^{n+1}	87
$H(t, x)$	92	$KR_{F2}(x)$	94	Ω	85
$H(\xi)$	118	Λ	84	Ω^*	85
$I(x:y)$	115	$l(x)$	84	$\overline{x}$	84
$J(\xi,:\eta)$	118	$\pi_1(z)$	84	$\subset$	85
$K_{F1}(x)$	88	$\pi_2(z)$	84	$\preccurlyeq$	85
$K(x)$	89	R	103	$\succcurlyeq$	85
$K(x \mid y)$	89	S	84	$\asymp$	85
$K_{F2}(x \mid y)$	89	S_ω	85		

Guide to the literature

The literature is referred to by section. The papers [5], [6], [11], [34] and textbooks [1], [37] seem especially useful to us for the relevant sections.

Preliminary remarks: [1]-[4]
§1 : [5]-[10]
§2 : [11]-[22], [33]
§3 : [8], [23]-[25]
§4 : [6], [7], [10], [22], [26]-[36] (the articles [27]-[32] are concerned with the concept of von Mises collective).
§5 : [5]-[7], [37]-[41] (the articles [37]-[40] are concerned with the classical concept of information).

In our paper we do not touch on questions connected with estimating the number of steps of an algorithm and the necessary size of memory, nor those connected with other aspects of the *complexity of calculation*. The reader who is interested in these questions can turn to the papers [42], [43] (where he will also find more references).

Our bibliography does not pretend to completeness. However, we have tried to include in it the principal publications supplementing the contents of our article.

References

[1] A.I. Mal'tsev, *Algoritmy i rekursivnye funktsii*, "Nauka", Moscow 1965. MR 34 # 2453.
Translation: Algorithms and recursive functions, Wolters-Noordhoff, Groningen 1970.

[2] V.A. Uspenskii, *Lektsii o vychislimykh funktsiyakh*, Fizmatgiz, Moscow 1960. MR 22 # 12043.
Translation: Lectures on computable functions, Pergamon, London 1966.

[3] S.C. Kleene, Introduction to meta mathematics, van Nostrand, Princeton 1952. MR 14-525.
Translation: *Vvedenie v metamatiku*, Inost. Lit. Moscow 1965. MR 19-2.

[4] A.A. Markov, Theory of algorithms, Trudy Mat. Inst. Steklov 42 (1954) MR 17-1038.
= Israel Programme for Scientific Translations, Jerusalem 1961. MR 24 # A 2527.

[5] A.N. Kolmogorov, Three approaches to the concept of the "amount of information", Problemy Peredachi Informatsii 1 : 1(1965), 3-11. MR 32 # 2273.

[6] A.N. Kolmogorov, Towards a logical foundation of information theory and probability theory, Problemy Peredachi Informatsii 5 : 3(1969), 3-7.

[7] A.N. Kolmogorov, Logical basis for information theory and probability theory, IEEE Trans. Information theory 14 (1968), 662-664.

[8] R.J. Solomonoff, A formal theory of inductive inference, Information and Control 7 (1964), 1-22.

[9] G.B. Marandzhyan, On certain properties of asymptotically optimal recursive function, Izv. Akad. Nauk Armyan. SSSR 4 (1969), 3-22.

[10] G.J. Chaitin, On the length of programs for computing finite binary sequences, I, II, J. Assoc. Comput. Math. 13 (1966), 547-570; 15 (1968)

[11] Ya.M. Barzdin', Complexity and the frequency solution of certain algorithmically unsolvable queuing problems, Preprint 1970.

[12] Ya.M. Barzdin', The complexity of programmes to determine whether natural numbers not greater than n belong to a recursively enumerable set, Dokl. Akad. Nauk SSSR 182 (1968), 1249-1252.
= Soviet Math. Dokl. 9 (1968), 1251-1254.

[13] Ya.M. Barzdin', On the relative frequency of solution of algorithmically ununsolvable queuing problems, Dokl. Akad. Nauk SSSR 191 (1970), 967-970.
= Soviet Math. Dokl. 11 (1970), 459-462.

[14] A.A. Markov, On normal algorithms associated with the computation of Boolean functions and predicates, Izv. Akad. Nauk SSSR Ser. Mat. 31 (1967), 161-208.

[15] A.A. Markov, On normal algorithms which compute Boolean functions, Dokl. Akad. Nauk SSSR 157 (1964), 262-264.
= Soviet Math. Dokl. 5 (1964), 922-924.

[16] M.I. Kanovich, On the decision complexity of algorithms, Dokl. Akad. Nauk SSSR 186 (1969), 1008-1009.
= Soviet Math. Dokl. 10 (1969), 700-701.

[17] M.I. Kanovich, On the complexity of enumeration and decision of predicates, Dokl. Akad. Nauk SSSR 190 (1970), 23-26.
= Soviet Math. Dokl. 11 (1970), 17-20.

[18] M.I. Kanovich and N.V. Petri, Some theorems on the complexity of normal algorithms and computations, Dokl. Akad. Nauk SSSR 184 (1969), 1275-1276.
= Soviet Math. Dokl. 10 (1969), 233-234.

[19] N.V. Petri, The complexity of algorithms and their operating time, Dokl. Akad. Nauk SSSR 186 (1969), 30-31.
= Soviet Math. Dokl. 10 (1969), 547-549.

[20] N.V. Petri, Algorithms connected with predicates and Boolean functions, Dokl. Akad. Nauk SSSR 185 (1969), 37-39.
= Soviet Math. Dokl. 10 (1969), 294-297.

[21] D.W. Loveland, A variant of the Kolmogorov notion of complexity, Preprint, 1970.

[22] P. Martin-Löf, On the variation of complexity of infinite binary sequences, Preprint 1970.

[23] Ya.M. Barzdin', On computability by probabilistic machines, Dokl. Akad. Nauk SSSR 189 (1969), 699-702.
= Soviet Math. Dokl. 10 (1969), 1464-1467.

[24] K. de Leeuw, E.F. Moore, C.E. Shannon and N. Shapiro, Computability by probabilistic machines, Automata studies, Princeton 1956. MR 18-104.
Translation: *Vychislimost' na veroyatnostnykh mashinakh*, Inost. Lit. Moscow 1956.

[25] V.N. Agafonov, On algorithms, frequency and randomness, Ph.D. dissertation, Novosibirsk 1970.

[26] A.N. Kolmogorov, *Osnovnye ponyatiya teorii veroyatnostei*, ONTI, Moscow 1936.
Translation: Foundations of the theory of probability, Chelsea, New York 1950.

[27] R. von Mises, Probability, statistics and truth, (2nd English edition), George Allen and Unwin, London 1957.
Translation: *Veroyatnost' i statistika*, Moscow-Leningrad 1930.

[28] A. Wald, Die Wiederspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung, Ergebnisse eines mathematischen Kolloquiums 8 (1937), 38-72.

[29] A. Church, On the concept of random sequence, Bull. Amer. Math. Soc. 46 (1940), 254-260.
[30] J. Ville, Étude critique de la notion de collectif, Gauthier-Villars, Paris 1939.
[31] A.N. Kolmogorov, On the tables of random numbers, Sankhya Ser. A 25 (1963), 369-376.
[32] D.W. Loveland, A new interpretation of the von Mises concept of random sequence, Z. Math. Logik Grundlagen Math. 12 (1966), 279-294.
[33] P. Martin-Löf, On the concept of a random sequence, Teor. veroyatnost i Primenen. 11 (1966), 198-200.
= Theory Probability Appl. 11 (1966), 177-179.
[34] P. Martin-Löf, The definition of random sequences, Information and Control 9 (1966), 602-619.
[35] P. Martin-Löf, Algorithms and random sequences, University of Erlangen, Germany, 1966.
[36] P.K. Schnorr, Eine neue Charakterisierung der Zufälligkeit von Folgen, preprint 1970.
[37] P. Billingsley, Ergodic theory and information, Wiley, New York 1965.
Translation: *Ergodicheskaya teoriya i informatsiya*, Mir, Moscow 1969.
[38] C.E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27 (1948), 379-423, 623-656.
[39] I.M. Gel' fand, A.N. Kolmogorov and A.M. Yaglom, Towards a general definition of the quantity of information, Dokl. Akad. Nauk SSSR 111 (1956), 745-748.
[40] A.N. Kolmogorov, A new metric invariant of transitive dynamical systems and automorphisms of a Lebesgue space, Dokl. Akad. Nauk SSSR 119 (1958), 861-864.
[41] A.N. Kolmogorov, Some theorems on algorithmic entropy and the algorithmic quantity of information, Uspekhi Mat. Nauk 23 : 2 (1968), 201.
[42] B.A. Trakhtenbrot, Complexity of algorithms and computations, Novosibirsk 1967.
[43] M. Blum, A machine-independent theory of the complexity of recursive functions, J. Assoc. Comput. Mach. 14 (1967), 322-337.

Received by the Editors,
7 August 1970.

Translated by S.M. Rudolfer.