INFORMATION AND CONTROL 9, 602–619 (1966)

The Definition of Random Sequences

PER MARTIN-LÖF

Institute of Mathematical Statistics, University of Stockholm, Stockholm, Sweden

Kolmogorov has defined the conditional complexity of an object y when the object x is already given to us as the minimal length of a binary program which by means of x computes y on a certain asymptotically optimal machine. On the basis of this definition he has proposed to consider those elements of a given large finite population to be random whose complexity is maximal. Almost all elements of the population have a complexity which is close to the maximal value.

In this paper it is shown that the random elements as defined by Kolmogorov possess all conceivable statistical properties of randomness. They can equivalently be considered as the elements which withstand a certain universal stochasticity test. The definition is extended to infinite binary sequences and it is shown that the non random sequences form a maximal constructive null set. Finally, the Kollektivs introduced by von Mises obtain a definition which seems to satisfy all intuitive requirements.

I. THE COMPLEXITY MEASURE OF KOLMOGOROV

Consider the set of all words over some finite alphabet. The length n of such a string $x = \xi_1\xi_2 \cdots \xi_n$ will be denoted by $l(x)$. Let A be an algorithm transforming finite binary sequences into words over some finite alphabet. We suppose that the algorithm concept has been made precise in one of the various equivalent ways that have been proposed, e.g. by means of the theory of partial recursive functions.

Following Kolmogorov we define the complexity of the element x with respect to the algorithm A as the length of the shortest program which computes it,

$$K_A(x) = \min_{A(p)=x} l(p).$$

If there is no such program, i.e. $A(p) \neq x$ for all binary strings p, we put $K_A(x) = +\infty$. This complexity measure depends in an essential way on the basic algorithm A. We almost get rid of this dependence by

means of the following theorem, proved by Kolmogorov and Solomonoff (1964).

There exists an algorithm A such that for any algorithm B

$$K_A(x) \leqq K_B(x) + c,$$

where c is a constant (dependent on A and B but not on x).

Such an algorithm is called asymptotically optimal by Kolmogorov and universal by Solomonoff. The complexity of x with respect to a fixed algorithm of this type we shall call simply the complexity of x and denote by $K(x)$.

In an analogous way we can introduce the concept of conditional complexity. To do this, let $p, x \to A(p, x) = y$ be an algorithm of two variables, where p is a finite binary sequence, called the program, x a string over some alphabet, and y a word over a possibly different alphabet. The quantity

$$K_A(y \mid x) = \min_{A(p,x)=y} l(p)$$

will be called the conditional complexity of y given x with respect to A.

There exists an algorithm A such that, for an arbitrary algorithm B,

$$K_A(y \mid x) \leqq K_B(y \mid x) + c,$$

where c is a constant (dependent on A and B but not on x and y).

A proof of this theorem, which is not more complicated than that of the previous one, was given by Kolmogorov (1965). Again we shall fix a universal algorithm, whose existence is guaranteed by the theorem, and write simply $K(y \mid x)$, speaking of the conditional complexity of y given x.

It is an immediate consequence of the theorem that there exists a constant c such that

$$K(\xi_1\xi_2 \cdots \xi_n \mid n) \leqq n + c$$

for every binary string $\xi_1\xi_2 \cdots \xi_n$. On the other hand, the number of sequences of length n for which

$$K(\xi_1\xi_2 \cdots \xi_n \mid n) \geqq n - c$$

is larger than $(1 - 2^{-c})2^n$, so that for large n the overwhelming majority of sequences $\xi_1\xi_2 \cdots \xi_n$ have a conditional complexity approximately equal to the maximal value n. Let us call these elements of maximal complexity random sequences. The thesis has been put forward by

Kolmogorov that this provides an adequate formalization of our intuitive notion of randomness.

II. A UNIVERSAL TEST FOR RANDOMNESS

In order to justify the proposed definition of randomness we have to show that the sequences, which are random in the stated sense, possess the various properties of stochasticity with which we are acquainted in the theory of probability. Assuming the binary alphabet to consist of the letters 0 and 1, the number of ones in $\xi_1\xi_2 \cdots \xi_n$ should be close to $n/2$, the number of zero runs to $n/4$, the number of occurrences of 0110 to $n/16$, and so on. It is not difficult to provide a proof in each of these cases, but the question arises whether it is possible to prove once and for all that the random sequences introduced possess, in some sense, all possible properties of stochasticity. Such a theorem should enable us to carry over automatically the various theorems of probability theory on random sequences. For example, with $s_n = \xi_1 + \xi_2 + \cdots + \xi_n$, we should be able to obtain a bound on $|2s_n - n|$ by means of $K(\xi_1\xi_2 \cdots \xi_n \mid n)$ and n, this bound being of the order of magnitude $\sqrt{n}$ when $K(\xi_1\xi_2 \cdots \xi_n \mid n)$ equals n approximately.

Let us borrow ideas from statistics. Consider a test for randomness, for example the one which rejects when the relative frequency of ones differs too much from $\frac{1}{2}$. Since we are always interested merely in the order of magnitude of the level of significance, we may restrict our attention to levels $\epsilon = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$. The particular test mentioned is given by the following prescription.

Reject the hypothesis of randomness on the level $\epsilon = 2^{-m}$ provided

$$|2s_n - n| > f(m, n).$$

Here f is determined by the requirement that the number of sequences of length n for which the inequality holds should be $\leqq 2^{n-m}$. Further, it should not be possible to diminish f without violating this condition.

Generally, a test is given by a prescription which, for every level of significance ϵ, tells us for what observations (in our case, binary strings) the hypothesis should be rejected. Taking $\epsilon = 2^{-m}$, $m = 1, 2, \cdots$, this amounts to saying that we have an effective description of the set

$$U \subseteq N \times X$$

(N denotes the set of natural numbers and X the set of all binary strings) of nested critical regions

$$U_m = \{x; m, x \in U\},$$

$$U_m \supseteq U_{m+1}, \qquad m = 1, 2, \cdots.$$

The condition that U_m be a critical region on the level $\epsilon = 2^{-m}$, amounts to requiring that the number of sequences of length n contained in U_m be $\leqq 2^{n-m}$.

Invoking the thesis of Church, we now formalize the fact that the family of critical regions is given by an explicit prescription by assuming the set U to be recursively enumerable. This is the weakest requirement we can imagine, and, in fact, all the tests of use in statistical practice are even of a much simpler type. In the following we shall, when speaking of a test, understand a recursively enumerable set U, interpreted as the family of critical regions, satisfying the restrictions above.

Having thus made precise the concept of a test, we are able to prove the following theorem, which, as will be shown below, could have been stated equivalently in terms of the conditional complexity measure and proved as a corollary of the second theorem of the previous section. Roughly speaking, it states that there exists a test, to be called universal, such that if a binary sequence is random with respect to that test, then it is random with respect to every conceivable test, neglecting a change in the level of significance.

There exists a test U such that, for every test V,

$$V_{m+c} \subseteq U_m, \qquad m = 1, 2, \cdots,$$

where c is a constant (dependent on U and V).

The proof is accomplished by first proving that the set of all tests is effectively enumerable.

There exists a recursively enumerable set $T \subseteq N \times N \times X$ such that U is a test if and only if

$$U = \{m, x; i, m, x \in T\}$$

for some $i = 1, 2, \cdots$.

It is well-known that the set of all recursively enumerable subsets of $N \times X$ is effectively enumerable. We exploit this fact by choosing a partial recursive function f of type $N \times N \to N \times X$ with the property that if it is defined for i, j, then $i, 1, i, 2, \cdots, i, j - 1$ likewise belong to the domain of definition. Further, a set in $N \times X$ is recursively enumerable if and only if it equals

$$\{f(i,j); j = 1, 2, \cdots\}$$

for some $i = 1, 2, \cdots$. The sets in this enumeration are now, if necessary, modified so that they all satisfy the conditions for a test. Remember that a recursively enumerable set $U \subseteq N \times X$ is defined to be a test if, firstly,

$$U_m \supseteq U_{m+1}, \qquad m = 1, 2, \cdots,$$

and, secondly, the number of elements of length n contained in U_m is $\leqq 2^{n-m}$ for all m and n. Fix an arbitrary $i = 1, 2, \cdots$. If $f(i,j)$ is undefined for all j, the corresponding recursively enumerable set is empty and hence trivially a test. Otherwise, calculate $f(i, 1) = m_1, x_1$. If the set of all m, x_1 for $m \leqq m_1$ satisfies the conditions for a test (in this case, $m_1 \leqq l(x_1)$), we include i, m, x_1 into T for all $m \leqq m_1$. Otherwise the section of T at i remains empty and the modification procedure is completed for this i. In the former case we proceed by calculating $f(i, 2) = m_2, x_2$ if defined and adding i, m, x_2 to T for all $m \leqq m_2$ provided the conditions for a test are not violated. If they are, the section of T at i is left unaffected by the last step and the modification is finished. It should now be evident how the construction is carried on. We note that the section of T at i is a test for every $i = 1, 2, \cdots$ which equals $\{f(i,j); j = 1, 2, \cdots\}$ provided this set already satisfies the definition of a test. The proof is finished.

The universal test U is obtained as the image of T under the mapping

$$i, m + i, x \to m, x.$$

For suppose that V is an arbitrary test. Then, for some i,

$$V = \{m, x; i, m, x \in T\},$$

so that

$$V_{m+i} = \{x; i, m + i, x \in T\} \subseteq \{x; m, x \in U\} = U_m$$

for all $m = 1, 2, \cdots$. We see that the constant c which figures in the theorem may be chosen as the Gödel number of the test V in the enumeration T.

As in statistical practice, it is convenient to introduce the critical level, the smallest level of significance on which the hypothesis is rejected. Since we have chosen to work with m instead of $\epsilon = 2^{-m}$, we introduce

$$m_U(x) = \max_{x \in U_m} m,$$

where the dependence on the particular test used is indicated by the subscript U. In order that $m_U(x)$ be defined for all x we define U_0 to be the set of all binary strings, so that

$$0 \leqq m_U(x) \leqq l(x)$$

for all x. In terms of the critical level the existence of a universal test can be stated thus. There exists a test U such that, for any test V, there is a constant c with the property that

$$m_V(x) \leqq m_U(x) + c$$

for all x. The critical level of x with respect to a fixed universal test we shall call simply the critical level of x and denote by $m(x)$. The relation to the complexity measure of Kolmogorov is given by the following theorem.

There exists a constant c such that

$$| l(x) - K(x \mid l(x)) - m(x)| \leqq c$$

for all binary strings x.

Define

$$V = \{m, x; K(x \mid l(x)) < l(x) - m\}$$

$$= \{m, x; (\exists p)(l(p) < l(x) - m \ \& \ A(p, l(x)) = x)\} \subseteq N \times X,$$

where A denotes the universal algorithm basic to the complexity measure. V is a test and

$$m_V(x) = l(x) - K(x \mid l(x)) - 1,$$

so that

$$l(x) - K(x \mid l(x)) \leqq m(x) + c$$

for some constant c.

To prove the inequality in the converse direction let U denote the universal test defining the critical level and choose a general recursive function f of type $N \to N \times X$ which enumerates U without repetitions. By means of f we construct the following algorithm from $X \times N$ to X. If $f(1) = m_1, x_1$, then

$$A(\underbrace{00 \cdots 00}_{l(x_1) - m_1}, l(x_1)) = x_1,$$

where the length of the string of zeros is $l(x_1) - m_1$. If $f(2) = m_2, x_2$ and $m_1, l(x_1) = m_2, l(x_2)$, then

$$A(\underbrace{00\cdots01}_{l(x_2)-m_2}, l(x_2)) = x_2,$$

otherwise

$$A(\underbrace{00\cdots00}_{l(x_2)-m_2}, l(x_2)) = x_2.$$

Since U is a test, the construction can be carried on without ambiguities. Evidently

$$K_A(x \mid l(x)) = l(x) - m(x),$$

so that

$$K(x \mid l(x)) \leqq l(x) - m(x) + c,$$

where c is a constant. The proof is finished.

Let us return to the concrete test considered in the beginning of this section. By means of the universal test we obtain the following inequality, holding for all binary strings $\xi_1\xi_2\cdots\xi_n$,

$$|2s_n - n| \leqq f(m(\xi_1\xi_2\cdots\xi_n) + c, n),$$

or, equivalently,

$$|2s_n - n| \leqq f(n - K(\xi_1\xi_2\cdots\xi_n \mid n) + c, n).$$

According to the theorem of de Moivre and Laplace,

$$\frac{1}{\sqrt{n}} f(m, n) \to \Phi^{-1}(1 - 2^{-m-1})$$

with

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy,$$

so that $|2s_n - n|$ is of the order of magnitude $\sqrt{n}$ provided $K(\xi_1\xi_2\cdots\xi_n \mid n)$ equals n approximately.

III. THE DEFINITION OF INFINITE RANDOM SEQUENCES

In the case of finite binary sequences the introduction of the universal test led to nothing but a useful reformulation of what could have been established by means of the complexity measure of Kolmogorov. We shall now see that by defining in a similar way a universal sequential test we obtain a natural definition of infinite random sequences. Such a definition has so far not been obtained by other methods.

Imagine a random device, such as the tossing of a coin, capable of delivering a potentially infinite binary sequence $\xi_1\xi_2 \cdots \xi_n \cdots$. To conform with our intuitive conception of randomness, such a sequence has to satisfy for example the law of large numbers,

$$\lim_{n\to\infty} \frac{s_n}{n} = \frac{1}{2},$$

or, requiring more, the law of the iterated logarithm,

$$\overline{\lim_{n\to\infty}} \frac{2s_n - n}{\sqrt{2n \log\log n}} = \pm 1.$$

In the measure theoretic probability theory this is motivated by proving that the set of all sequences violating the law has measure zero. By definition this means that to every $\epsilon > 0$ there exists an open covering $\mathfrak{U}$ of the set such that

$$\mu(\mathfrak{U}) \leqq \epsilon.$$

Here μ denotes the usual measure with respect to which all coordinates are independent and take on the values 0 and 1 with probability $\frac{1}{2}$. Let $\mathfrak{I}(\xi_1\xi_2 \cdots \xi_n)$ denote the set of all infinite sequences beginning with $\xi_1\xi_2 \cdots \xi_n$. Then, instead of $\mathfrak{U}$, we may just as well consider the set

$$U = \{x; \mathfrak{I}(x) \subseteq \mathfrak{U}\} \subseteq X.$$

Note that, conversely,

$$\mathfrak{U} = \bigcup_{x\in U} \mathfrak{I}(x)$$

if and only if $\mathfrak{U}$ is open. Further, U has the property that it contains all possible extensions of any of its elements, y being an extension of x, in symbols $y \geqq x$, if the string y begins with x. In other words, U may be regarded as the critical region of a sequential test on the level ϵ. The definition of a null set may hence be stated in statistical terms as follows. For every $\epsilon > 0$ there exists a sequential test on that level which rejects all sequences of the set.

We can now argue just as in the previous section. Any sequential test of present or future use in statistics is given by an explicit prescription, which, for every level of significance $\epsilon = \frac{1}{2}, \frac{1}{4}, \cdots$, tells us for what sequences the hypothesis is to be rejected. Equivalently, when proving the law of large numbers or some other theorem involving the words almost surely, we actually construct an open covering of measure $\leqq \epsilon$ for

arbitrarily small ϵ, which, without restriction of generality, we may take to be of the form 2^{-m}, $m = 1, 2, \cdots$. These statements are made precise by assuming that the family of critical regions (open coverings)

$$U \subseteq N \times X$$

is recursively enumerable. U has to satisfy the following restrictions. If $m, x \in U$, then so does n, y for all $n \leqq m$ and $y \geqq x$. Further, the number of sequences of length n contained in

$$U_m = \{x;\ m, x \in U\}$$

is $\leqq 2^{n-m}$ for all m and n.

Again we can prove the key theorem to the effect that the set of all sequential tests (open coverings) is effectively enumerable.

There exists a recursively enumerable set $T \subseteq N \times N \times X$ such that U is a sequential test if and only if

$$U = \{m, x;\ i, m, x \in T\}$$

for some $i = 1, 2, \cdots$.

The proof differs only negligibly from that of the previous section. We choose the partial recursive function f just as before, fix an arbitrary $i = 1, 2, \cdots$ and calculate $f(i, 1) = m_1, x_1$ if defined. Provided we do not violate the conditions connected with the level of significance (in this case, $m_1 \leqq l(x_1)$) we include into T i, m, x for all $m \leqq m_1$ and $x \geqq x_1$. Otherwise the section of T at i remains empty. If we have not finished already, we continue by trying to calculate $f(i, 2) = m_2, x_2$ and including i, m, x for all $m \leqq m_2$ and $x \geqq x_2$. These indications should suffice.

There exists a universal sequential test U such that, for any sequential test V,

$$V_{m+c} \subseteq U_m, \qquad m = 1, 2, \cdots,$$

where c is a constant (dependent on U and V).

Again U is obtained as the image of T under the mapping

$$i, m + i, x \to m, x.$$

It is readily verified that U is a sequential test satisfying the conditions of the theorem.

The critical level

$$m_U(x) = \max_{x \in U_m} m$$

with respect to a sequential test U satisfies not only

$$0 \leqq m_U(x) \leqq l(x)$$

but also

$$m_U(x) \leqq m_U(y)$$

for all $x \leqq y$. Consequently, we can introduce the critical level of an infinite sequence $\xi_1\xi_2 \cdots \xi_n \cdots$

$$m_U(\xi_1\xi_2 \cdots \xi_n \cdots) = \lim_{n\to\infty} m_U(\xi_1\xi_2 \cdots \xi_n),$$

$$0 \leqq m_U(\xi_1\xi_2 \cdots \xi_n \cdots) \leqq +\infty.$$

Having fixed a universal test U, we shall drop the index U and speak simply of the (sequential) critical level.

An infinite binary sequence $\xi_1\xi_2 \cdots \xi_n \cdots$ is called a random sequence provided

$$m(\xi_1\xi_2 \cdots \xi_n \cdots) < +\infty.$$

Note that this definition does not depend on the choice of the universal test with respect to which the critical level is defined.

Almost all infinite binary sequences are random sequences.

Introduce the open sets

$$\mathfrak{U}_m = \bigcup_{x\in U_m} \mathfrak{I}(x), \qquad m = 1, 2, \cdots.$$

Since U is a sequential test,

$$\mathfrak{U}_1 \supseteq \mathfrak{U}_2 \supseteq \cdots$$

and

$$\mu(\mathfrak{U}_m) \leqq 2^{-m}, \qquad m = 1, 2, \cdots$$

The set of all nonrandom sequences is precisely the null set

$$\bigcap_{m=1}^{\infty} \mathfrak{U}_m$$

provided U was chosen universal.

Let us make another reformulation, this time in the spirit of constructive analysis. An open set $\mathfrak{U}$ of infinite binary sequences is called constructively open if $\{x;\ \mathfrak{I}(x) \subseteq \mathfrak{U}\}$ is recursively enumerable. $\mathfrak{U}_1, \mathfrak{U}_2, \cdots$ is a constructive sequence of constructively open sets provided $\{m, x;\ \mathfrak{I}(x) \subseteq \mathfrak{U}_m\}$ is recursively enumerable. α is defined to be a constructive null set if

$$\mathfrak{A} \subseteq \mathfrak{U}_m, \qquad m = 1, 2, \cdots,$$

where $\mathfrak{U}_1, \mathfrak{U}_2, \cdots$ is a constructive sequence of constructively open sets such that

$$\mu(\mathfrak{U}_m) \to 0$$

constructively fast as $m \to \infty$. By this we understand that $\mu(\mathfrak{U}_m) \leqq 2^{-k}$ for all $m \geqq h(k)$, where h is a general recursive function. In this terminology we can say that the set of all nonrandom sequences form a maximal constructive null set, i.e., a constructive null set $\mathfrak{A}$ with the remarkable property that any constructive null set $\mathfrak{B}$ is contained in it. For let $\mathfrak{B}$ be an arbitrary constructive null set and $\mathfrak{V}_1, \mathfrak{V}_2, \cdots$ the associated coverings. Without restriction of generality we may assume that

$$\mathfrak{V}_1 \supseteq \mathfrak{V}_2 \supseteq \cdots,$$

$$\mu(\mathfrak{V}_m) \leqq 2^{-m},$$

so that

$$V = \{m, x; \mathfrak{z}(x) \subseteq \mathfrak{V}_m\}$$

is a sequential test. According to the definition of a universal sequential test U

$$V_{m+c} \subseteq U_m, \qquad m = 1, 2, \cdots,$$

for some constant c. Consequently,

$$\mathfrak{B} \subseteq \bigcap_{m=1}^{\infty} \mathfrak{V}_m = \bigcap_{m=1}^{\infty} \mathfrak{V}_{m+c} \subseteq \bigcap_{m=1}^{\infty} \mathfrak{U}_m = \mathfrak{A},$$

where, as before,

$$\mathfrak{U}_m = \bigcup_{x \in U_m} \mathfrak{z}(x), \qquad m = 1, 2, \cdots.$$

IV. RANDOM SEQUENCES WITH RESPECT TO AN ARBITRARY COMPUTABLE PROBABILITY DISTRIBUTION

So far we have introduced random sequences that were to represent the result of tossing a perfect coin. We shall now see that in a similar way we can introduce finite and infinite sequences which are random with respect to an arbitrary computable probability distribution.

Let $p(x)$ denote the probability of the binary string x (or, better, the conditional probability of x given its length). The conditions to be satisfied by p are as usual

$$p(x) \geqq 0, \qquad \sum_{l(x)=n} p(x) = 1$$

for all n. By the computability of p we understand that p is a general recursive function which for every x calculates a Gödel number of the computable real number $p(x)$.

In the case of a random device giving out sequentially a potentially infinite binary sequence, the probability $p(x)$ that the first n digits equal $x = \xi_1\xi_2 \cdots \xi_n$ must satisfy

$$p(x) \geqq 0, \qquad p(\,) = 1,$$
$$p(x) = p(x0) + p(x1)$$

for all x. The computability of p is defined as before.

A test for p is a recursively enumerable set

$$U \subseteq N \times X$$

with the usual property that

$$U_1 \supseteq U_2 \supseteq \cdots,$$

the condition on the level being

$$\sum_{x \in U_m, l(x)=n} p(x) < 2^{-m}$$

for all m and n. The choice of strict inequality is due to the fact that if a and b are computable real numbers such that $a < b$, we will get to know this sooner or later by calculating the successive approximations to a and b. This does not hold, in general, when $<$ is changed to $\leqq$.

A sequential test for a sequential computable probability distribution is defined in the same way except for one additional condition. With x the critical region U_m has to contain all $y \geqq x$, $m = 1, 2, \cdots$.

Using the technique that has been demonstrated twice already, we can prove the effective enumerability of all (sequential) tests for a certain (sequential) computable probability distribution and hence the existence of a corresponding universal (sequential) test. The critical level is introduced and, in the sequential case, extended to infinite sequences,

$$0 \leqq m(\xi_1\xi_2 \cdots \xi_n \cdots) = \lim_{n\to\infty} m(\xi_1\xi_2 \cdots \xi_n) \leqq +\infty.$$

Finite binary strings x are random with respect to the computable probability distribution considered, provided the critical level $m(x)$ is

low. In the infinite case the dependence on the choice of the universal sequential test disappears, $\xi_1\xi_2 \cdots \xi_n \cdots$ being by definition random if

$$m(\xi_1\xi_2 \cdots \xi_n \cdots) < +\infty.$$

The set of all nonrandom sequences is precisely

$$\bigcap_{m=1}^{\infty} \mathfrak{U}_m,$$

where U denotes the universal sequential test and

$$\mathfrak{U}_m = \bigcup_{x \in U_m} \mathfrak{I}(x).$$

Letting π denote the measure (in the usual sense) obtained by extending the computable probability distribution p, the set of all random sequences has measure one with respect to π. It would be natural to call it the constructive support of π.

V. FINITE BERNOULLI SEQUENCES

For an arbitrary binary string $\xi_1\xi_2 \cdots \xi_n$ with $s_n = \xi_1 + \xi_2 + \cdots + \xi_n$ put

$$p(\xi_1\xi_2 \cdots \xi_n) = \theta^{s_n}(1 - \theta)^{n-s_n},$$

where $0 \leqq \theta \leqq 1$. If θ is a computable real number, this defines a computable probability distribution and the results of the preceding section can be applied to obtain a definition of finite and infinite Bernoulli sequences associated with a computable success probability. These are precisely the Bernoulli sequences that can be produced by a computing machine with access to a table of random numbers as defined in Sections II and III, and so we have met exactly the needs of the Monte Carlo theory. However, we cannot be satisfied with this as a mathematical description of the sequences obtained, e.g., by tossing an imperfect coin. Indeed, there seems to be no reason whatsoever to assume that such a success probability, thought of as a physical constant associated with the coin, is a computable real number.

We shall, instead, define Bernoulli sequences without using any measure theoretic concepts, by merely requiring that the successes be located at random. In other words, a Bernoulli sequence is a sequence whose only regularities are given by the frequencies of successes and failures. This is connected with the statistical concept of sufficiency. Indeed, the

success and failure frequencies form a sufficient statistic for the class of all Bernoulli distributions.

A test for the Bernoulli property or, simply, a Bernoulli test is given by a recursively enumerable set

$$U \subseteq N \times X$$

such that

$$U_1 \supseteq U_2 \supseteq \cdots \supseteq U_m \supseteq \cdots .$$

Further, the number of sequences with s_n ones and $n - s_n$ zeros contained in U_m should be

$$\leqq 2^{-m} \binom{n}{s_n}$$

for all m, n and s_n. Thus, the test is carried out as a conditional test. Now everything can be carried out just as before.

There exists a recursively enumerable set $T \subseteq N \times N \times X$ such that U is a Bernoulli test if and only if

$$U = \{m, x; i, m, x \in T\}$$

for some $i = 1, 2, \cdots$.

There exists a universal Bernoulli test U such that if V is an arbitrary such test,

$$V_{m+c} \subseteq U_m , \qquad m = 1, 2, \cdots ,$$

for some constant c.

Finite Bernoulli sequences are those strings whose critical level (with respect to a fixed universal test),

$$m(x) = \max_{x \in U_m} m,$$

is low. Again we could have reached this definition equivalently by means of the complexity measure of Kolmogorov. In terms of that concept the Bernoulli sequences are defined by requiring the conditional complexity, given the frequencies of zeros and ones, to be maximal, i.e., approximately equal to

$$\log \binom{n}{s_n} .$$

Here and in the sequel the logarithm is taken to the base two. Note that

$$K(\xi_1 \xi_2 \cdots \xi_n \mid s_n, n - s_n) \leqq \log \binom{n}{s_n} + c,$$

where c is a constant.

There exists a constant c such that

$$\left| \log \binom{n}{s_n} - K(\xi_1 \xi_2 \cdots \xi_n \mid s_n, n - s_n) - m(\xi_1 \xi_2 \cdots \xi_n) \right| \leqq c$$

for all binary strings $\xi_1 \xi_2 \cdots \xi_n$.

The proof so closely parallels that of the corresponding theorem of Section II, that there is no need to give it in detail.

Let us make a slight but illuminating digression. The interpretation of a probability is currently (e.g., in the Grundlagen by Kolmogorov) governed not only by the clause that the relative frequency in a large number of repetitions of the experiment should be close to it, but also by the following somewhat obscure additional clause. If the probability is very small, we should be practically sure that the event does not occur in a single trial. In the present formalism we can show that if $\xi_1 \xi_2 \cdots \xi_n$ is a Bernoulli sequence with a very low relative success frequency s_n/n, then, necessarily, $\xi_1 = 0$, so that the event cannot have occurred in the first trial. In other words, the assumption that a success occurred already in the first trial implies substantial regularities in the sequence.

There exists a constant c such that

$$m(\xi_1 \xi_2 \cdots \xi_n) \leqq \log \frac{n}{s_n} - c$$

implies $\xi_1 = 0$.

Construct the test which rejects on the level $\epsilon = 2^{-m}$ when $\xi_1 = 1$ and $s_n/n \leqq 2^{-m}$. Then the number of rejected sequences of length n with success frequency s_n equals

$$\binom{n-1}{s_n - 1} = \frac{s_n}{n} \binom{n}{s_n} \leqq 2^{-m} \binom{n}{s_n},$$

so that the definition is legitimate. Comparison with the universal test yields the theorem.

VI. INFINITE BERNOULLI SEQUENCES

The definition of infinite Bernoulli sequences is now straightforward.

We note that these are precisely the sequences for which von Mises introduced the term Kollektiv. In our case the Merkmalraum consists merely of two elements, but the extension to an arbitrary finite number is trivial.

A sequential Bernoulli test is a recursively enumerable set

$$U \subseteq N \times X$$

which together with m, x includes n, y for all $n \leqq m$ and $y \geqq x$. Further, the number of strings of length n with s_n successes contained in $U_m = \{x; m, x \in U\}$ should be

$$\leqq 2^{-m} \binom{n}{s_n}$$

for all m, n and s_n .

There exists a recursively enumerable set $T \subseteq N \times N \times X$, such that the sequential Bernoulli tests are precisely the sets

$$\{m, x; i, m, x \in T\}, \qquad i = 1, 2, \cdots .$$

Maybe it is worth while pointing out the following simple fact which is needed in the proof. Let A be a set of strings of length n and let a_i denote the number of strings in A containing i successes. We suppose that

$$a_i \leqq 2^{-m} \binom{n}{i}, \qquad i = 0, 1, \cdots, n.$$

Let B be the set of all strings of length $n + 1$ whose initial segments belong to A, and define b_j in analogy with a_i . Then,

$$b_j = a_{j-1} + a_j \leqq 2^{-m} \left(\binom{n}{j-1} + \binom{n}{j} \right) = 2^{-m} \binom{n+1}{j},$$

$j = 0, 1, \cdots n + 1$. Using this the proof is not more complicated than that of Section III. Taking again the image of T under the mapping

$$i, m + i, x \to m, x$$

we obtain a universal test.

There exists a universal sequential Bernoulli test U such that, for any sequential Bernoulli test V,

$$V_{m+c} \subseteq U_m, \qquad m = 1, 2, \cdots ,$$

where c is a constant.

Allowing infinite values, the critical level with respect to a sequential Bernoulli test is extended to infinite sequences. Bernoulli sequences (Kollektivs, in the terminology of von Mises) are defined by the requirement that the critical level (with respect to a universal test) be finite,

$$m(\xi_1\xi_2 \cdots \xi_m \cdots) < +\infty.$$

Let π_θ denote the measure over the space of infinite binary sequences with respect to which all coordinates are independent and Bernoulli distributed with success probability θ, $0 \leqq \theta \leqq 1$.

The set of Bernoulli sequences has measure one with respect to π_θ for all $0 \leqq \theta \leqq 1$.

As before, put

$$\mathfrak{U}_m = \bigcup_{x \in U_m} \mathfrak{I}(x)$$

so that

$$\sum_{x \in U_m, l(x)=n} \theta^{s_n}(1-\theta)^{n-s_n} \uparrow \pi_\theta(\mathfrak{U}_m)$$

as $n \to \infty$. But

$$\sum_{x \in U_m, l(x)=n} \theta^{s_n}(1-\theta)^{n-s_n} \leqq 2^{-m} \sum_{s_n=0}^{n} \binom{n}{s_n} \theta^{s_n}(1-\theta)^{n-s_n} = 2^{-m},$$

and hence

$$\pi_\theta(\mathfrak{U}_m) \leqq 2^{-m}, \qquad m = 1, 2, \cdots,$$

$$\pi_\theta\left(\bigcap_{m=1}^{\infty} \mathfrak{U}_m\right) = 0.$$

Note that the set of Bernoulli sequences is the complement of

$$\bigcap_{m=1}^{\infty} \mathfrak{U}_m.$$

The aim of the present paper has only been to give the basic definitions. It is, however, too difficult to resist the temptation of proving two important properties of Bernoulli sequences. Remember that our definition is a kind of irregularity condition in that we require the successes to be located at random, no restriction being laid upon the frequencies. It is a remarkable fact that the existence of the limit of the relative frequency as the number of trials grows beyond all bounds is a consequence of this

irregularity condition. Recall that in the tentative definition of von Mises the convergence of the relative frequencies is introduced as a postulate, which is supplemented by a kind of irregularity condition.

Let $\xi_1\xi_2 \cdots \xi_n \cdots$ be an infinite Bernoulli sequence. Then the relative frequency s_n/n converges as $n \to \infty$.

For an arbitrary rational $\epsilon > 0$ we construct the test which rejects on the level 2^{-m} provided

$$\left| \frac{s_i}{i} - \frac{s_j}{j} \right| > \epsilon$$

for some $i, j \geqq h(m)$, where h is a suitable nondecreasing general recursive function, an explicit definition of which we could evidently write down with some effort. A comparison with the universal test completes the proof.

Note that, by the law of large numbers, all real numbers θ (not only computable ones) occur as limit frequencies,

$$\lim_{n\to\infty} \frac{s_n}{n} = \theta, \qquad 0 \leqq \theta \leqq 1.$$

We finally state the analogue of the last theorem of the previous section, the idea of the proof being the same.

The limit frequency cannot vanish,

$$\lim_{n\to\infty} \frac{s_n}{n} = 0,$$

unless $\xi_n = 0$ for all n.

This theorem is important since, in the case of an experiment with an arbitrary finite number of outcomes, it allows us to reduce the sample space by excluding those outcomes whose limit frequencies equal zero. More suggestively, an event with vanishing limit frequency is actually impossible. This contrasts sharply with the conception of von Mises, who explicitly stated that the opposite might occur. It seems as if he strained his seldom failing intuition on this point in order not to conflict with his somewhat arbitrary definition of randomness.

RECEIVED: April 1, 1966

REFERENCES

KOLMOGOROV, A. N., (1965), Tri podhoda k opredeleniju ponjatija "količestvo informacii." *Problemy peredači informacii* **1**, 3–11.

SOLOMONOFF, R. J., (1964), A formal theory of inductive inference. Part I. *Inform. Control* **7**, 1–22.