ON LOGICAL FOUNDATIONS OF PROBABILITY THEORY[*)]

A.N.KOLMOGOROV

In everyday language we call random these phenomena where we cannot find a regularity allowing us to predict precisely their results. Generally speaking there is no ground to believe that a random phenomenon should possess any definite probability. Therefore, we should have distinguished between randomness proper (as absence of any regularity) and stochastic randomness (which is the subject of the probability theory).

There emerges a problem of finding the reasons for applicability of the mathematical theory of probability to the phenomena of the real world. As for me, personally, my first experience to tackle the problem in this direction was the paper [1]. (It was published in an edition of a methodological nature).

Since randomness is defined as absence of regularity, we should primarily specify the concept of regularity. The natural means of such a specification is the theory of algorithms and recursive functions; the first attempt of its application in probability theory was that made by Church [2].

The aim of my report is to acquaint the audience with this range of concepts in the first approximation.

Paying a tribute to the tradition we shall begin with the classic definition of the probability as the ratio of the number of favourable outcomes to the total number of outcomes

$$P=\frac{m}{n},$$

where n is the total number of all possible outcomes (of one trial) m is the number of favourable outcomes. This definition actually reduces the problem of calculating the probability to the combinatorial problems.

However, this definition cannot be applied in many practical situations. This is what gave an impetus to the emergence of the so-called statistical definition of probability

$$P \approx \frac{\mu}{N}, \qquad (*)$$

where N is the total number of trials which is assumed to be sufficiently large, μ is the number of successes.This definition,its initial form, is, strictly speaking, not a mathematical one. For this reason the formula (*) contains the symbol of an approximate equality.

[*)] The report is recorded by Novikov A.A., Zvonkin A.K.,Shen' A.

The first attempts to make the definition (*) sound more exact were made by R.von Mises. But before we start describing his approach, let us discuss (from the viewpoint of the classic definition of probability) the question of why we so often observe the stability of frequences in natural phenomena.

Consider the set of all 0-1-sequences of length n containing exactly m units and assume all such sequences equally probable.

Let a certain way be given of dividing any 0-1-sequence of length n into two subsequences. Then for each sequence it is possible to compare the frequences of units in both subsequences having calculated the difference

$$\left| \frac{\mu_1}{n_1} - \frac{\mu_2}{n_2} \right|$$

(n_1 and n_2 are lengths of subsequences, μ_1 and μ_2 are the numbers of units in them, $n_1 + n_2 = n$, $\mu_1 + \mu_2 = m$). We would like to expect the difference to be small almost surely in the sense that

$$P_{class} \left\{ \left| \frac{\mu_1}{n_1} - \frac{\mu_2}{n_2} \right| < \varepsilon \right\} \longrightarrow 1 \qquad as\ n_1, n_2 \longrightarrow \infty .$$

Certainly, to make the assertion true, it is necessary to narrow down the class of possible rules of selecting subsequences (in particular, having forbidden the following rule: to select in one subsequence only zeros, and into the other only units).

The paper [3] contains necessary specifications to the concept of the admissible rule of selecting a subsequence based on the ideas by Mises. The concept of admissible rule plays a crucial part in Mises frequency approach to the concept of probability. According to Mises, the infinite sequence $x_1, x_2, \ldots$ of zeros and units is called a Bernoulli one if:

(1) there exists the limit $P = \lim\limits_{n \to \infty} \frac{1}{n} \sum\limits_{i \leq n} x_i$;

(2) the limit remains constant if we pass from the entire sequence to its subsequence obtained by means of the admissible rule of selection:

$$\lim_{m \to \infty} \frac{1}{m} \sum_{j \leq m} x_{n_j} = P .$$

As to the rule of selection, Mises here gave only a general outline and examples. As a matter of fact, they are reduced to the fact that the selection of the next chosen member of the subsequence must not depend on its value, but must be defined by the values of the already selected members. This is, of course, not an exact definition ,but no such definition could be expected to arise since the concept itself of the rule had no strict mathematical analogue at that time. The situation changed essentially when there appeared the concepts of an algorithm and a recursive function. With their help, Church [2] specified Mises' definitions. In the abovementioned paper [3] a class of selection algorithms was proposed broader than that by Church. According to [3] , the rule of selection is given by means of an algorithm (or, if you like, by Turing machine). Selection of the next member of the subsequence takes place in the following way. The input information consists of the finite sequence of the numbers $n_1, n_2, \dots, n_k$ and values $x_{n_1}, x_{n_2}, \dots, x_{n_k}$ of the members of the initial sequence. The output of the algorithm is, firstly, the number n_{k+1} of the next scanned element $x_{n_{k+1}}$ (this number must coincide with none of those $n_1, \dots, n_k$; as to the order of the numbers $n_1, \dots, n_k$, no restrictions are laid to it); secondly, the indication whether $x_{n_{k+1}}$ is selected only to be scanned or the algorithm decided to include $x_{n_{k+1}}$ into the sequence selected.

On the next step of the algorithm's work its input consists already of a longer sequence of numbers $n_1, \dots, n_{k+1}$ and values $x_{n_1}, \dots, x_{n_{k+1}}$; the algorithm naturally starts its work from the empty set.

Expansion, as compared to [2] , consists in the fact that the order of members in the selected subsequence should not obligatorily coincide with their order in the initial subsequence. Another, even more important difference of [3] from the papers by Church and Mises consists in a strictly finite nature of the entire conception and in introducing the quantitative evaluation of the frequencies stability mentioned above.

Passage to the finite sequences unavoidably requires the introduction of the restrictions to the complexity of the selection algorithm. Exact definition of the complexity of a finite object and pattern of applying it to the probability theory foundations were proposed in the papers [3], [6].

Results obtained under the frequency approach and the complexity one are compared in Shen' [4] .

Now let us return to the initial idea that "randomness" consists in the absence of "regularity" and show in what way the concept of complexity of a finite object allows us to attach exact meaning to it. A lot of papers have been devoted to the concept of complexity; they are majorly divided into two groups: papers on the complexity of calculations and those on the complexity of definitions. We will deal with the latter. Below is given the definition of complexity from [6] . We define conditional complexity of a constructive object with respect to a certain algorithm A under the condition that the constructive object Y is known. To be more precise, define the conditional complexity $K_A(X/Y)$ of the object X , Y being known, as a length of the minimal program by means of which algorithm A can obtain X from Y:

$$K_A(X/Y) = min\{\ell(p) | A(p,Y) = X\}.$$

Here $\ell(p)$ is a length of 0-1-sequences regarded as a program. There exists an "optimal" algorithm A such that for any algorithm A_1 there exists such a constant C that for all X and Y

$$K_A(X/Y) \le K_{A_1}(X/Y) + C.$$

If A_1 and A_2 are optimal algorithms, the complexity functions given by them differ no more than by a constant (independent of X and Y).

Now we can define the concept of a "random", or , to be more precise, Δ-random object in a given finite set M (here Δ is a number). Namely, we shall say that X M is Δ-random in M if

$$K_A(X/Y) \geqslant log_2|M| - \Delta,$$

where M denotes the number of elements in M . We shall call random in M the Δ-random objects in M , Δ being comparatively small.Thus we receive the definition of a random finite object which can be regarded as a final one.

If we take as M the set D_n of all the 0-1-sequences of the length n , we come to the condition : $K_A(X/D_n) \geqslant n - \Delta$.

It may be proved that for the sequence having this property, Δ being sufficiently small, there is, in particular, fulfilled the property of frequencies stability in the selection of subsequences.

Thus, requirements to randomness formulated by Mises, prove to be a particular case of our requirements. Further results in this direction may be found in the papers [5] , [7] - [12].

REFERENCES

[1] Kolmogorov A.N. (1956),Theory of probability .(In the book "Mathematics, its contents, methods and meaning ", Moscow).

[2] Church A.(1940), On the concept of a random sequence,Bulletin of Amer. Math. Soc.,46 ,N 2, 130-135.

[3] Kolmogorov A.N. (1963),On tables of random numbers, Sankhya, Ser.A, v.25, part 4, 369-376.

[4] Shen' A. (1982),Frequency approach to the definition of notion of a random sequence, "Semiotika and Informatika",N 18,p.14-42, Moscow, VINITI,(in Russian).

[5] Loveland D. (1966), A new interpretation of the von Mises' concept of random sequence, Z.fur Math. Logic und Grundlagen der Math., Bd.12.,H.4, 279-294.

[6] Kolmogorov A.N. (1969), Towards a logical foundation of information theory and probability theory, Problems in the transmission of information, v.5,N3, p.3-7 (in Russian).

[7] Solomonoff R.J. (1964), A formal theory of inductive inference, Inform. and Control, v.7, N1, p. 1-22.

[8] Martin-Löf P. (1966), Algorithms and Random Sequences, Univ. of Erlangen.

[9] Martin-Löf P. (1966), The definition of Random Sequences, Inform. and Control,v.9, 6, 602-619.

[10] Vjugin V.V, (1981), Algorithmic entropy (complexity) of finite objects andits application to the definition of randomness and quantity of information, "Semiotika and Informatika", 16, p.14-43 (in Russian).

[11] Chaitin G. (1966), On the length of programs for computing finite binary sequences, J.ASM, v.13, N4, p.547-569.

[12] Zvonkin A.K.,Levin L.A. (1970),The complexity of finite objects and the development of concepts of information and randomness by means of the theory algorithms, Russian Math. Surveys, v.25, N6, p.83-124 .

[13] Kolmogorov A.N. (1983),Combinatorial basis of information theory and probability theory, Report on International Mathematical Congress, Nice 1970, Russian Math. Surveys, v.38, N4, (in Russian).

Moscow State University
117234 Moscow USSR