Albert N. SHIRYAEV

Steklov Mathematical Institute,
Lomonosov Moscow State University

Probabilistic and algorithmic approaches to the concept of randomness

e-mail: albertsh@mi.ras.ru

§ 1. INTRODUCTION

1.1. Our presentation focuses on the notion of

RANDOMNESS

or, more exactly, on the question

how one can define formally what is

INDIVIDUAL RANDOM SEQUENCE.

In the paper “On logical foundation of Probability Theory” (Fourth USSR–Japan Symposium, 1982) A. N. Kolmogorov wrote:

> 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 definitive probability. Therefore, we should have distinguished between
>
> **proper** randomness
>
> (as absence of any regularity)
>
> and
>
> **stochastic** randomness
>
> (which is the subject of the probability theory).

Let us take, for example, the *finite* Bernoulli sequence

$$\boxed{(\mathrm{I}_{10})}: \qquad \mathbf{0\,1\,1\,1\,0\,1\,0\,0\,1\,0}$$

or the *infinite* Bernoulli sequence

$$\boxed{(\mathrm{I}_{00})}: \qquad \mathbf{0\,1\,1\,1\,0\,1\,0\,0\,1\,0\,1\,1}\ldots = \boxed{(\mathrm{I}_{10})}\,\mathbf{1\,1}\ldots,$$

which are formed by “fair” tossing of a regular coin ($\mathbf{1}$ = Head, $\mathbf{0}$ = Tail).

Having get these sequences, we shall be inclined to qualify them as

“RANDOM”,

since they seem to display **no regularity** in consecution of zeroes and units.

But if we get any of the sequences

$\boxed{(\mathrm{II}_{10})}$: **1 1 1 1 1 1 1 1 1 1**, which consists of ten units,

$\boxed{(\mathrm{III}_{10})}$: **1 0 1 0 1 0 1 0 1 0**, which consists of alternating units and zeroes,

then our intuition will hardly allow us to reckon them as “random”.

However, from the probabilistic point of view, each of the sequences

$$(\mathrm{I}_{10}), \quad (\mathrm{II}_{10}) \quad \text{and} \quad (\mathrm{III}_{10})$$

has $\boxed{\text{the same probability } (1/2)^{10}}$.

(We consider a symmetric Bernoulli scheme, where it is assumed that at each step the coin is tossed independently and the probabilities of getting 1 or 0 are equal to 1/2.)

Our aim is to review in a compact form the main approaches—which have been proposed by now—to the following problem:

Which sequence can be naturally named "random"?

The ideal result would be to "split" the set of all binary sequences into two sets, random and nonrandom sequences.

For what follows, it is important to emphasize again that

the sequences that we consider are not arbitrary but those obtained from probabilistic experiments which are described by the symmetric Bernoulli scheme.

Thus, we start with an assumption that we are in the framework of the Kolmogorov axiomatics, which assumes given a measurable space of outcomes $(\Omega, \mathcal{F})$ with certain probability measure P.

A priori it would be natural to think that

the problem of “splitting” sequences into
“random” and “nonrandom”

(within one or another definition of randomness which would conform to our informal intuition)

can be solved in the framework of
the **theory of probability**

However, the above example of sequences (I_{10}), (II_{10}) and (III_{10}), which have the same probability, shows that it is scarcely possible.

One cannot say that probability theory is not at all able to give answers to the questions of “splitting”. The actual situation is that the Kolmogorov axiomatics of probability theory is designed in a way which allows one to establish one or another property only P-a.s. In other words, the theory of probability answers the question:

Is a certain property fulfilled?

only for “overwhelming majority” of objects (for example, binary sequences), without determining which concrete individual objects belong to this “majority”.

It turned out that solution of the problem whether a concrete individual object is ‘‘random’’ or ‘‘nonrandom’’ can be helped by addressing to a discipline, which seems to be very distant from the theory of probability, namely, the *theory of algorithms*.

Here we should give a very important warning: In principle, it is hardly possible to draw a clear-cut distinction between ‘‘randomness’’ and ‘‘nonrandomness’’ for *finite* sequences.

As concerns the case of *infinite* sequences, we will see that the theory of algorithms provides the reasonable definitions of an *infinite random sequence* which fit well our intuition.

It is well known that in Probability Theory for

infinite sequences

there are many results about the validity of the “overwhelming majority” properties formulated as properties fulfilled

almost surely.

For example, consider the Bernoulli probability space:

$$(\Omega, \mathcal{F}, \mathsf{P}) \equiv \Big(\{-1,1\}^{\infty},\ \mathcal{B}(\{-1,1\}^{\infty}),\ \mathsf{P}_{\mathsf{Bern}}\Big).$$

Set, for $x = (x_1, x_2, \ldots)$

$$S_n(x) = x_1 + \cdots + x_n, \qquad n \geq 1.$$

The classical results of Probability Theory claim that for the random walk $(S_n(x))_{n\geq 1}$ the following properties hold $\mathbf{P}_{\mathbf{Bern}}$**-almost surely**:

- $\displaystyle \lim_{n\to\infty} \frac{S_n(x)}{n} = 0$ (strong law of large numbers);
- $\displaystyle \limsup_{n\to\infty} \frac{S_n(x)}{\sqrt{n}} = +\infty, \qquad \liminf_{n\to\infty} \frac{S_n(x)}{\sqrt{n}} = -\infty;$
- $\displaystyle \frac{S_n(x)}{\sqrt{n}\log n} \to 0, \qquad n \to \infty;$
- $\displaystyle \limsup_{n} \frac{S_n(x)}{\sqrt{2n\log\log n}} = 1$ (law of the iterated logarithm).

It is remarkable that (**V. G. Vovk, A. Shen**)

> the algorithmic approach to the validity of the formulated properties allows one to describe the **individual** sequences $x = (x_1, x_2, \ldots) \in \Omega$ for which these statements do hold.

1.2. Richard von Mises was the first to consider (in 1919) the notion

‘infinite **random** sequence’.

His intention was in fact to build the probability theory assuming the notion of ‘infinite random sequence’ as a basis.

Note that the axiomatics of probability theory which was proposed by Kolmogorov (1933) and is generally accepted nowadays is based on a different object, namely,

probability distribution.

In the very beginning (§ 2) of his monograph “Grundbegriffe der Wahrscheinlichkeitsrechnung” (Springer, Berlin, 1933) Kolmogorov, emphasizing the importance of the von Mises approach for the probability theory, wrote:

> Exposing necessary precondition for applicability of probability theory to the world of real events the author follows to a considerable extent the conclusions of von Mises, cf., in particular: R. von Mises. Vorlesungen aus dem Gebiete der angewiesen Mathematik. Bd. 1: Wahrscheinlichkeitsrechnung. Leipzig u. Wien, Fr. Deuticke, 1931 (cf. pp. 21-27, Section “Das Verhältnis der Theorie zur Erfahrungswelt”).

Von Mises did not give an accurate formal mathematical definition of the notion ‘*random* sequence’, contenting himself with reference to intuitive ideas of

- “irregularity of their construction”,
- “unpredictability of their subsequent values by the preceding ones”,
- impossibility to construct winning strategies over the sequences produced in casino.

However, the idea itself to define a “random” sequence generated a large cycle of works which developed various natural approaches to definition of the concept of “randomness”.

As we already mentioned, all these approaches are based on a concept of ‘algorithms’, which seems to be alien to the theory of probability. However, it is the theory of algorithms that gave the possibility to specialize the uncertain von Mises notion of ‘admissible selection rules’ (‘admissible place selection’) that he used for definition of the notion ‘infinite random sequence’.

By von Mises, an admissible place selection is a procedure for selecting a subsequence of a given sequence $x = (x_1, x_2, \ldots)$ in such a way that the decision to select a term x_n does not depend on the value x_n.

Now one can distinguish four main approaches to the definition of the concept of ‘infinite random sequence’. These approaches are based on and determined by the following four properties that our intuition demands from sequences which we call ‘random’:

STABILITY of FREQUENCIES or stochasticity	Mises, Wald Church, Kolmogorov Loveland
TYPICALITY (belonging to a set with effective measure 1)	Martin-Lŏf Levin Schnorr
COMPLEX STRUCTURE, or CHAOTICITY	Kolmogorov Levin Schnorr
NONPREDICTABILITY	Ville, Uspensky

After V. A. Uspensky * , each of these properties represents

> “its own algoritmic physiognomy of randomness, and each of them can—with more or less ground—pretend to a role of an accurate mathematical definition of the concept of randomness”.

* V. A. Uspensky, *Four algorithmic physiognomies of randomness* (Russian), Matematicheskoe prosveshchenie, **10**, MCCME, Moscow, 2006, 71–108.

§ 2. TOWARDS the HISTORY of COMING-TO-BE of PROBABILITY THEORY

To clarify the questions connected with notions of ‘probability’ and ‘randomness’, it is worth recalling now the main stages of the probability theory coming-to-be as a mathematical discipline.

Both intuitive ideas on *randomness* and beginnings of reasoning of different kinds about possible *chances* (in religious practice, settlement of controversies, predictions, . . .) trace their roots back to ancient days when the manifestations of randomness were believed to be divine apparition which is beyond the reach of a human mind.

Archeological finds say about existence of “random tools”, namely hexahedral dice (*astragalus*[*]), already in ancient times[**].

In the Renaissance (end of the XIV century – beginning of the XVII century), we find traces of more or less serious discussions (generally, of a *philosophical* character) on “probabilistic” reasoning:

Fra Luca Pacioli	(1445–1517(?))
Celio Calcagnini	(1479–1541)
Nicola Fontana Tartaglea	(1500–1557)

[*] **astragalus** is a heel bone of Artiodactyla; it has such a form that, when tossed up, it can fall on one of four (different) sides, since the other two have a rounded form.

[**] in the period of the First Dynasty in Egypt (c. 3500 before Christ), then in Ancient Greece and Ancient Rome, where they were used in primitive games.

One of the first to analyze **mathematically** the gaming chances was

Gerolamo CARDANO (1501–1576)

who solved the cubic equation and is broadly known as inventor of “cardan shaft”. In his manuscript (c. 1525)—published only in 1663 under the title *Liber de Ludo Aleæ* (*Book on games of chance*)—he launched an idea of **combinations** which opened a convenient way of describing the set of all outcomes and the set of favorable outcomes.

The period briefly portrayed above is commonly referred to as *prehistory* of the theory of probability. When studying the questions of the history of probability theory one usually distinguish the following five stages:

PREHISTORY

1st PERIOD (XVII century – beginning of the XVIII century)

2nd PERIOD (XVIII century – beginning of the XIX century)

3rd PERIOD (the latter half of the XIX century)

4th PERIOD (beginning and middle of the XX century)

FIRST PERIOD (XVII century – beginning of the XVIII century)

This period is commonly associated with the birth of "calculus of probabilities", and its starting point is fastened with the correspondence (1654) between **Blaise Pascal** (1623–1662) and **Pierre de Fermat** (1601–1665).

In 1657, **Christianus Huygens** (1629–1695) publishes his book

De Ratiociniis in Aleæ Ludo (*"Games of chance"*).

This book was received warmly by contemporary mathematicians and remained—for nearly half a century—a unique introduction to the theory of probability.

A central figure of the considered period is certainly represented by

JACOB (Jakob, James, Jacques) **BERNOULLI** (1654–1705)

who is credited to introduce into the science the notion **‘probability of an event’**. J. Bernoulli was the first

- to consider *infinite* sequences of iterated trials and

- to put a question about the limiting behavior of the frequencies of appearing of one or another event in these trials, which was a cardinally new (“nonfinitistic”) idea in probabilistic considerations, which were restricted at that time to methods of the elementary arithmetics and simplest techniques of combinatorics; this setup led Bernoulli to the

  law of large numbers,

  which bears now his name (“Ars Conjectandi”, 1713).

SECOND PERIOD (XVIII century – beginning of the XIX century)

This period is tied up with such names as

Pierre-Rémond de **Montmort**	(1678–1719)
Abraham De **Moivre**	(1667–1754)
Thomas **Bayes**	(1702–1761)
Pierre Simon de **Laplace**	(1749–1827)
Carl Friedrich **Gauss**	(1777–1855)
Siméon Denis **Poisson**	(1781–1840)

In his book Essai d'Analyse sur les Jeux de Hasard (Essay of Analysis in Games of Chance), **Montmort** (1708) pays a special attention to development of methods of calculation in various games.

In the books Doctrine of Chances (1718) and Miscellanea Analytica Supplementum (1730) **Moivre** gives definitions of such notions as

- *independence* of events,
- *expectation*,
- *conditional probability*.

The name of Moivre is mostly reputed in connection with the normal approximation of the binomial distribution, which—at the suggestion of George Pólya (1920)—is called now a

CENTRAL LIMIT THEOREM.

A work by **T. Bayes** "An Essay Towards Solving a Problem in the Doctrine of Chances" (1763) provided

$$\boxed{\textbf{BAYES' FORMULA}},$$

a rule of conversion of a priori probabilities into a posteriori probabilities after a given event.

On the heels of J. Bernoulli, **Laplace** held to the "classical" definition of probability (in the case of finitely many possible outcomes with equal probabilities).

However, “nonclassical” probability arose as early as in this period. For example, in the Bayes considerations a posteriori probabilities can be nonequal.

Nevertheless, the objects known now as Gaussian law and Poisson’s law were considered only as approximations and their essence as probability distributions (in the modern interpretation of this term) was not realized yet.

The aforesaid shows that the framework of the “classical” (finitistic) probability theory began to restrict severely its development. In this period, probability theory failed abstract mathematical constructions and it was qualified just as applied mathematics.

THIRD PERIOD (the latter half of the XIX century)

The main locality where general problems of probability theory were developed at that time was **St.-PETERSBOURG**, and an essential contribution in extension and deepening of the whole system of probability was made by

P. L. Chebyshev	(1821–1894)
A. A. Markov	(1856–1922)
A. M. Lyapunov	(1857–1918)

It is to their works that one owes

the refusal to restrict oneself within the case of
“classical” probability.

CHEBYSHEV appreciated with the utmost clarity the role of the notion of random variable; he developed a new method of the proof of limit theorems, namely the **method of moments** which was perfected later by

MARKOV who introduced also a fundamentally new concept, namely a scheme of dependent variables
which form a **“Markov chain”**.

An unexpected step in finding general conditions for the “de Moivre–Laplace theorem” to hold was made by

LYAPUNOV who elaborated the **method of characteristic functions** that allowed him to prove the “Central Limit Theorem” under assumption that the summands are independent and their moments of order $2+\delta$, $\delta>0$, are finite.

In **WESTERN EUROPE** in the latter half of the XIX century, the interest in the theory of probability began to grow rapidly thanks to discovering of its profound connections with

pure mathematics,

statistical physics,

and **mathematical statistics**
which began to develop quickly at that time.

Let us cite just a few names:

Henri **Poincaré** (1854–1912)	recurrent motions in dynamic systems
Hugo **Gyldén** (1841–1896)	problems of stability of planets and the probabilistic number theory
James Clerk **Maxwell** (1831–1879)	Maxwell’s distribution for molecular velocities
Ludwig **Boltzmann** (1844–1906)	*time averages* and *ergodicity hypothesis*
Josiah Willard **Gibbs** (1839–1903)	notion of an ensemble and the “Gibbs distribution”

The following discoveries were of a great importance for all the subsequent development of probability theory as well as for the deeper understanding of the role of probabilistic approaches and concepts:

- a phenomenon discovered in 1827 by **Robert Brown** (1773–1858) and named afterwards a **'Brownian motion'**.

  Qualitative explanation and quantitative description of the Brownian motion were proposed later on by **Albert Einstein** (1879–1955) and **Marian Smoluchowski** (1872–1917).

- a phenomenon of **radioactive decay**, discovered in 1896 by **Antoine-Henri Becquerel** (1852–1908) when investigating the properties of uranium.

  This phenomenon found its explanation in the framework of **quantum mechanics**, whose creation relates to 1920es.

FOURTH PERIOD (the beginning and middle of the XX century)

Connections of probability theory with pure mathematics which were revealed by the end of XIX century, led to the setup—by **David Hilbert** (1862–1943), in his programming lecture on 8 August 1900 at the Second Mathematical Congress in Paris—of a problem of

***MATHEMATIZATION* of Probability Theory**.

Among the well-known problems launched by Hilbert, the *sixth* * was formulated as the problem of

> axiomatization of those physical disciplines in which mathematics plays a dominating role.

Among such disciplines D. Hilbert ranked

probability theory and **mechanics**.

* The *first* problem concerned the continuum hypothesis.

The fourth period in the history of the coming-to-be of probability theory is a period of creation of its logical grounds and of its becoming a mathematical discipline.

Shortly after the lecture of D. Hilbert, some attempts to construct the *mathematical* theory of probability using elements of the theory of sets and the theory of measure were made.

In 1904 **R. Lämmel**, to describe the set of outcomes, turned to the **theory of sets**, but the notion of probability remained on intuitive level and was associated with volume, area, length, ...

In 1907 **U. Broggi**, in his thesis advised by D. Hilbert, appealed to the Borel–Lebesgue **theory of measure**, but the definition of the notion itself of (finitely additive) probability needed calling for "relative measures", "relative frequencies" (in the simplest cases) and for some artificial limiting procedures (in the general case).

In 1917 **S. N. Bernstein** proposed a system of axioms based on the notion of ***qualitative* comparison of events** according to the degree (greater or smaller) of their likelihood. As regard the numerical value of probability it arose as a derivative notion.

In later 1920s and early 1930s, **B. de Finetti** developed a very similar approach, which was based on *subjective* qualitative judgements ("**knowledge system of a subject**").

In 1919, R. von Mises proposed the so-called

FREQUENTIST APPROACH

[also referred to as **statistical**
or **empirical**]

to the logical grounds of probability theory, basing on the idea that

> probabilistic concepts can be applied only to the so-called **“collectives”**, i. e., *individual infinite ordered sequences* which have a certain property of “randomness” of their formation.

Now we came near the main theme of our lecture, a survey of the von Mises approach to the notion of “randomness” and its subsequent developments.

§ 3. Von Mises' frequentist probabilities

Richard von Mises (19.04.1883–14.07.1953) was an applied mathematician, far-famed for his works in mechanics and especially in hydrodynamics, theory of flights. (In Harvard University he was professor of aerodynamics and applied mathematics.) His contribution to formation of grounds of the theory of probability (1919) * is very important.

Von Mises, first of all, was interested in the applicability of probability theory to real world phenomena. This is why he considered the theory of probability as a doctrine of mass phenomena and, consequently, reckoned it as a natural science, discipline which is determined by the specificity of "mass phenomena". (Compare with physics, biology, . . . , which have a certain mathematical specificity.)

* **Grundlagen der Wahrscheinlichkeitsrechnung**, *Math. Z.* 5 (**1919**), 52–99.

As a natural science, the doctrine of “mass phenomena” can be studied by various methods (including mathematical ones), but within the frame of its subsect. *

Von Mises realized that to ground the theory of probability one need a certain idealizations of the subject; he suggested that the study of probability is intricately related with the study of random sequences.

More exactly, the scheme adopted by von Mises can be depicted in the following way.

* Those who follows the Kolmogorov axiomatics can think of the theory of probability as of a mathematical discipline which is a part of a general theory of sets and functions.

Given a sample space (“Merkmalraum”) M of points (“labels”), we assume that we are able to make infinitely many trials which will give a sequence $x = (x_1, x_2, \ldots)$, where x_n is an outcome (with values in M) in nth trial. Let A be a subset of the phase space M and let

$$\nu_n(A; (x_k)_{k \leq n}) = \frac{1}{n} \sum_{k=1}^{n} I_A(x_k)$$

be frequency of appearance of an “event” A in n trials which give the sequence of labels $(x_1, \ldots, x_n)$.

After Mises, an infinite sequence $x = (x_1, x_2, \ldots)$ is called

a COLLECTIVE,

if the following two postulates are fulfilled:

(A) the limit $\boxed{\lim_{n\to\infty} \nu_n(A; (x_k)_{k\le n}) \ \ (= P(A;x))}$ exists;

(B) the limit $\boxed{\lim_{n\to\infty} \nu_n(A; (x'_k)_{k\le n}) \ \ (= P(A;x'))}$ exists for all subsequences $(x'_k)_{k\ge1} = (x'_1, x'_2, \ldots)$ which are obtained from the sequence $x = (x_1, x_2, \ldots)$ by means of *any* **“admissible”** choice of elements $x'_k = x_{n_k}$, $n_1 < n_2 < \cdots$, of the sequence $x = (x_1, x_2, \ldots)$.

It is assumed that for all “admissible” sets A and all “admissible” selection rules the limits $P(A;x')$ must coincide with the limit $P(A;x)$.

EXAMPLE of extracting an **admissible subsequence** x' from the sequence

$$x = 0\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ 1\ 0 \ldots$$

One reads the sequence x from left to right and mark (by $\boxed{0\,1}$) the “words” 0 1:

$$x = 0\,\boxed{0\,1}\,\boxed{0\,1}\,\boxed{0\,1}\,1\,1\,\boxed{0\,1}\,1\,\boxed{0\,1}\,0 \ldots$$

$\Downarrow$

The sequence x' is formed by figures which go just after the string 01:

$$x = 0\,\boxed{0\,1}\,\boxed{\mathbf{0}\,1}\,\boxed{\mathbf{0}\,1}\,\mathbf{1}\,1\,\boxed{0\,1}\,\mathbf{1}\,\boxed{0\,1}\,\mathbf{0} \ldots$$

$\Downarrow$

Admissible subsequence is

$$x' = \mathbf{0\ 0\ 1\ 1\ 0} \ldots$$

Another example of an admissible selection rule: the sequence x' is composed of elements $x_{i_1}, x_{i_2}, \ldots$ of x whose numbers i_k are primes.

In the von Mises approach, the second postulate (B), which is aimed to reflect the idea of "randomness", of absence of "regularity" in the structure of collectives, is of particular importance.

The aforesaid shows that, constructing probability on a sample space X, von Mises proceed from the assertion that

> this probability can be defined only in connection with existence of collectives which have "random origin" (according to postulate (B)).

The postulate (B) of “randomness” provoked the serious criticism of the whole (frequentist) von Mises approach to the ground of probability theory, first of all because von Mises did not give a formal definition of “admissible” selection rules. (To justify the very existence of collective he referred to the existence of gambling houses, to impossibility of construction of winning strategies against the “random” sequences $x = (x_1, x_2, \ldots)$ produced in casinos.)

Somehow or other, the great merit of the von Mises frequentist theory of probability theory was that postulate (B) formulated by von Mises stimulated the investigation of the problem: Which of infinite sequences meet our idea of “randomness”?

§ 4. STABILITY of FREQUENCIES (or STOCHASTICITY)

As was already mentioned, Mises did not give a formal-logic definition of a notion of "random" sequence, since he did not specialize the "admissible rules of choice" (or "admissible place selection") and thus it was not clear which subsequences must satisfy the postulate (B).

$\langle$Recall again that, by von Mises, an admissible place selection is a procedure for selecting a subsequence of a given sequence $x = (x_1, x_2, \dots)$ in such a way that the decision to select a term x_n does not depend on the value x_n.$\rangle$

Moreover, some set-theoretic aspects (additivity, countable additivity) for the "probabilities" $P(A, x)$ were not clarified.

F. Hausdorff in his letter to G. Pólya (January 1920) expressed his doubt in *existence* of "collectives" with property of invariance of frequencies.

One of the first who tried to give a logically consistent formulation of what are “admissible” rules of choice of subsequences and prove that the class of collectives is nonempty was

A. WALD * : His idea consisted in

constructing “admissible” sequences by means of **functions** $W = W(x^{(n)})$, defined on the chains $x^{(n)} = (x_1, \ldots, x_n)$, $n \geq 1$, assuming that each of these functions takes one of two values, 1 или 0.

(It is convenient to introduce an “empty” chain $x^{(0)}$ and define the value $W(x^{(0)})$ on this chain to be 1 or 0.)

* **Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung**, *Ergebnisse eines mathematischen Kolloquiums*, **8**, 38–72; **1937** .

Construction of a subsequence which is “W-ADMISSIBLE” after WALD:

if $W(x^{(0)}) = 1$, then x_1 is included in the subsequence,
if $W(x^{(0)}) = 0$, then x_1 is not included;

$W(x^{(1)}) \equiv W(x_1)$;

if $W(x_1) = 1$, then x_2 is included in the subsequence,
if $W(x_1) = 0$, then x_2 is not included;

$\cdots\cdots\cdots$

That is,

the “admissible” (or “W-admissible”) sequence is

$$\boldsymbol{x_{n(1)}, x_{n(2)}, \ldots} \quad \textbf{where} \quad \begin{aligned} &\boldsymbol{n(1) = \min\{k \geq 0\colon W(x^{(k)}) = 1\},} \\ &\boldsymbol{n(2) = \min\{k > n(1)\colon W(x^{(k)}) = 1\},} \\ &\boldsymbol{\ldots} \end{aligned}$$

The basic result of A. Wald is a proof of the **NONEMPTINESS of the class of “collectives”**. Namely,

> Let S be an arbitrary countable system of admissible rules each of which is determined by its own collection of functions W. Then there exist infinitely many sequences which satisfy postulates (A) и (B).

! However, one cannot guarantee that the limiting set functions $P(A;x)$ are countably additive, and without this property one cannot consider some important questions, for example, the question whether the law of the iterated logarithm holds.

The analysis of Wald’s admissible sequences shows that they can be sufficiently “regular” and not sufficiently disorderly though our intuition suggests that “random” sequences are sufficiently disordered.

In **1940** **Alonzo Church** ("On a concept of a random sequence"), one of creators of the theory of algorithms, launched an idea that subsequences to be included in a collective should be selected **effectively**. Thus he suggested that the methods of choosing a subsequence should be restricted to those we can actually perform. For the realization of this idea, he proposed to attract the newly-formulated concept of recursiveness. That concept, according to the well-known Church theses, was meant as a mathematically precise formulation of the **algorithm** * compatibility.

* "An **ALGORITHM** is an finite sequence of instructions, an explicit, step-by-step procedure for solving a problem, often used for calculation and data processing. It is formally a type of effective method in which a list of well-defined series of successive states, eventually terminating in an end-state, is defined. A function is called **algorithmically computable** if there exists an algorithm which computes it.

.

The accurate definition of the notion ‘computable function’ was given by **A. Church** (1936) who proposed to identify the computable function having natural arguments and values with the notion ‘general recursive function’. (He gave also first example of the noncomputable function.) In 1936, **E. Post** and **A. Turing** gave the first specification of the notion ‘algorithm’ using the terms of the idealized computer machines. The most general definition of the notion of algorithm was proposed in 1953 by **A. N. Kolmogorov**. He proved that this very general definition is reduced to an algorithm of calculation of values of a partially recursive function.

Examples of algorithms: rules of addition, subtraction, multiplication, long division. A prototypical example of an “algorithm” is Euclid’s algorithm of determining the greatest common divisor of two integers which are > 1. Other examples are dynamic programming and linear programming algorithms.

Since the class of algorithm is countable, the class of computable functions is also countable and, consequently, the class $R(\text{MWCh})$ of sequences which are random after Mises–Wald–**Church** is nonempty: $R(\text{MWCh}) \neq \varnothing$.

Unfortunately, the class $R(\text{MWCh})$ turned out to be still too wide: **J. Ville** (1939) constructed a sequence which is random according to the Mises–Wald–**Church** definition but has too much regularity to be called random. For example, for this sequence the law of iterated logarithm is not fulfilled, although this law is natural for random sequences.

In 1966 **D. Loveland** noticed that the class $R(\text{MWCh})$ of sequences which are "random" after Mises–Wald–**Church** contains sequences which, after a certain computable permutation of their entries, are no longer in the class $R(\text{MWCh})$.

Taking into account these circumstances, Kolmogorov in 1963 proposed

> a narrowing of the class R(MWCh) at the expense of enlargement of the set of “admissible selection rules” (tests) which satisfy the von Mises postulate (B). *

* Note that D. Loveland in 1966 came to the same construction.

The **Church's subsequences** (obtained by means of one or another computable function) were of the form

$\boxed{x_{n(1)}, x_{n(2)}, \cdots}$ where $n(i) < n(j)$ for $i < j$,

while the **Kolmogorov admissible (generalized) subsequences** (obtained by means of **two** computable functions ψ and φ) were of the form

$\boxed{x_{\varphi(1)}, x_{\varphi(2)}, \cdots}$ where $\varphi(i) \neq \varphi(j)$, if $i \neq j$. (φ)

The Kolmogorov procedure consists in two steps: first one construct (by means of a computable function ψ) a generalized sequence

$\boxed{x_{\psi(1)}, x_{\psi(2)}, \cdots}$ (with $\psi(i) \neq \psi(j)$, if $i \neq j$), (ψ)

then from this sequence one extracts (by means of a computable function φ, this latter procedure repeats the Church one) a subsequence (φ).

The class R(K) obtained in such a way satisfies

$$R(\mathsf{K}) \subset R(\mathsf{MWCh}).$$

Unfortunately, this class R(K) also turned out too wide to be taken as a “class R(**?**)” of truly “random” sequences. The cause is that in the class R(K) there exist sequences in each initial segment of which the number of units exceeds the number of zeroes, the fact contradicting * both

- our intuition (about “uniformity” of appearance of units and zeroes in “random” sequences obtained in the symmetrical Bernoulli scheme) and
- a law of probability theory valid for this scheme bearing the name of “recurrency law”.

* Exceeding of the number of units over the number of zeroes should correspond to the case $p > 1/2$ in the Bernoulli scheme with probability of “success” equal to p; as is well known, in this case the random walk is nonrecurrent.

Thus, we have the chain of inclusions

$$\boxed{R(\mathbf{?}) \subset R(\mathsf{K}) \subset R(\mathsf{MWCh})}$$

($R(\mathbf{?})$ is a class of “truly random” sequences obtained by the von Mises “admissible” selection rules).

The results stated above exhaust in the main the progress obtained in the way of constructing sequences with stability of frequencies. It is worth emphasizing that

> **a truly random class** $R(\mathbf{?})$ of “admissible” sequences with stability of frequencies, for which the main laws of probability theory would remain true,
>
> **IS NOT DETERMINED YET**.

§ 5. TYPICALITY (belonging to a set with effective measure 1)

To describe a different (namely, "typical") approach to the notion of randomness, which was initiated by **P. Martin-Lőf** (**1966**), recall first the Borel strong law of large numbers.

Let $\Omega = [0,1)$. Denote by $\mathcal{B}$ the Borel system of subsets of Ω and by P the Lebesgue measure on $[0,1)$.

Consider a binary record $\boxed{x = 0.x_1 x_2 \ldots}$ of $x \in \Omega$ (with infinitely many zeroes) and define random variables $\xi_1(x), \xi_2(x), \ldots$ by letting $\xi_n(x) = x_n$. For any $n \geq 1$ and all $b_1, b_2, \ldots$ taking values 0 and 1,

$$\{\omega\colon \xi_1(x) = b_1, \ldots, \xi_n(x) = b_n\} = \left\{ \sum_{i=1}^{n} \frac{b_i}{2^i} \leq x < \sum_{i=1}^{n} \frac{b_i}{2^i} + \frac{1}{2^n} \right\},$$

so the measure P of this set equals $1/2^n$. Thus, $\xi_1(x), \xi_2(x), \ldots$ are i.i.d. random variables with $\mathrm{P}(\xi_i = 0) = \mathrm{P}(\xi_i = 1) = 1/2$.

BORFL STRONG LAW OF LARGE NUMBERS says that

almost all numbers $x = 0.x_1x_2\ldots$ from the interval $[0, 1)$ are normal in a sense that **with probability 1** the part of zeroes (and the par of units) in a binary record of x tends to $1/2$, i.e., for "most" x's (more exactly, P-a.s.)

$$\boxed{\frac{1}{n}\sum_{k=1}^{n} I(x_k = 1) \longrightarrow \frac{1}{2}}. \qquad (*)$$

Thus, the P-measure of $x = 0.x_1x_2\ldots$ such that the limit

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} I(x_k = 1)$$

either does not exist, or does not equal $1/2$, is 0.

Such P-null sets are called **negligible**. Our intuition suggests that

> if, say, a set $U \subseteq \Omega$ is negligible (i.e., P$(U) = 0$), then all elements $x = 0.x_1x_2\ldots$ of this set (in other words, all sequences $(x_1, x_2, \ldots)$) should be proclaimed **“NONTYPICAL”** (i.e., not belonging to “majority”), since the property $(*)$ fails.

But such sets U of “untypical” x’s are, generally speaking, “large in number” and therefore

the set of “untypical” x’s should be taken as a maximal set which contains all sets U with P$(U) = 0$; in other words,

$$\boxed{\text{set of “nontypicality”} = \bigcup U}$$

where the sum (which is, generally speaking, not countable) is taken over all U such that P$(U) = 0$.

HOWEVER, it is well known that it can occur that the P-measure of the set $\bigcup U$ is well-defined and P$(\bigcup U) = 1$. Thus,

> **the set of all "untypical" x** $= 0.x_1x_2\ldots$ (in other words, of sequences $(x_1, x_2, \ldots)$)—which seems to be a natural candidate for the nomination as "a set of 'nonrandomness' "—**has the measure 1**.

It is clear that the set of all "typical" x's is $\bigcap \overline{U}$ $(= \overline{\bigcup U})$ so that

> **the set of "typical" x's**, which we want to consider as the set of "randomness", **is generally P-null**—the fact which **contradicts both the strong law of large numbers and the common sense**.

A way to overcome the difficulty of defining “typical” sequences which we would like to proclaim “random” was proposed, as we already said, by **P. Martin-Lőf** who specialized the notion of a

‘negligible set’

(i.e., a set with the measure 0); namely, he introduced (in an algorithmic way) a new notion of

‘EFFECTIVELY negligible set’.

It is this notion that allowed the natural definition of “typical sequences” which we want to declare “random”.

Recall that in the theory of measure a set U is called negligible or P-null, if P$(U) = 0$. In the case of the space Ω of sequences it is convenient to use the following (equivalent) definition of negligible sets.

$\Omega := \{x\colon x = x_1x_2\ldots\}$ is a set of ALL binary (i.e., $x_i = 0$ or 1) sequences (of infinite words).

$\boldsymbol{\Xi} := \{\xi\colon \xi = x_1x_2\ldots x_{|\xi|}\}$ is a set of FINITE binary words ξ, with lengths $|\xi|$ taking values in $\{1, 2, \ldots\}$.

Ω_ξ is a set of infinite binary words with initial fragment $\xi \in \boldsymbol{\Xi}$; in other words, Ω_ξ is a cylindrical set with the "base" ξ.

DEFINITION I. A set $U \subseteq \Omega$ is called

negligible (with respect to the measure P),

if for any integer $m \geq 1$ one can find a sequence $\xi_1, \xi_2, \ldots$ of binary words (from $\boldsymbol{\Xi}$) such that

$$U \subseteq \bigcup_n \Omega_{\xi_n}, \quad \text{where} \quad \sum_n \mathrm{P}(\Omega_{\xi_n}) = \sum_n 2^{-|\xi_n|} < \frac{1}{m}.$$

REMARK.

Each one-point set U, which consists of a single sequence $x \in \Omega$, is evidently negligible, since the initial fragments of length n have probability 2^{-n} and it suffices to choose an n such that $2^{-n} < 1/m$.

The following notion of effectively negligible set, which is needed for our purposes, specializes the above definition of a negligible set.

DEFINITION II. A set $U^* \subseteq \Omega$ is called

EFFECTIVELY negligible

(with respect to the measure P),

if for any integer $m \geq 1$ there exists an m-effectively computable sequence $\xi_1^*, \xi_2^*, \ldots$ of binary words (from $\boldsymbol{\Xi}$) such that

$$U^* \subseteq \bigcup_n \Omega_{\xi_n^*}, \quad \text{where} \quad \sum_n \mathrm{P}(\Omega_{\xi_n^*}) = \sum_n 2^{-|\xi_n^*|} < \frac{1}{m}.$$

In Definition II, the qualifier

“m-effectively computable”

is defined **by means of ALGORITHMIC notions**:

We say that a sequence of binary words

- $\xi_1, \xi_2, \ldots$ is **algorithmically computable**, if there exists an algorithm which calculates each ξ_n by its number n.

- $\xi_1^*, \xi_2^*, \ldots$ is **m-effectively computable**, if there exists an algorithm which, starting from a number m, elaborates another algorithm (program) which creates an algorithmically computable sequence.

DEFINITION I. A set $U \subseteq \Omega$ is called

negligible (w.r.t. the measure P),

if for any integer $m \geq 1$ one can find a sequence $\xi_1, \xi_2, \ldots$ of binary words (from Ξ) such that

$$U \subseteq \bigcup_n \Omega_{\xi_n}, \quad \text{where} \quad \sum_n \mathrm{P}(\Omega_{\xi_n}) = \sum_n 2^{-|\xi_n|} < \frac{1}{m}.$$

DEFINITION II. A set $U^* \subseteq \Omega$ is called

EFFECTIVELY negligible (w.r.t. the measure P),

if for any integer $m \geq 1$ there exists an $\underline{m\text{-effectively}}$ $\underline{\text{computable}}$ sequence $\xi_1^*, \xi_2^*, \ldots$ of binary words (from Ξ) such that

$$U^* \subseteq \bigcup_n \Omega_{\xi_n^*}, \quad \text{where} \quad \sum_n \mathrm{P}(\Omega_{\xi_n^*}) = \sum_n 2^{-|\xi_n^*|} < \frac{1}{m}.$$

The importance of the notion of an effectively negligible set is revealed by the following result by **P. Martin-Lőf**:

TEOPEMA. There exists an effectively negligible set which contains ALL effectively negligible sets.

Thus, if the completion of an effectively negligible set is said to be 'effectively large', then *the intersection of all effectively large sets is again an effectively large set and has the (effective) measure* 1.

This is the above intersection that one proclaims to be a set of **"TYPICAL" sequences**, denote by $R(\mathrm{T})$ and calls a set of

RANDOM SEQUENCES after MARTIN-LÖF.

It is worth our attention to observe that

for “typical” sequences (i.e., the sequences of the class $R(\mathsf{T})$) the basic laws of probability theory, including the law of iterated logarithm, are fulfilled.

To the chain of inclusions

$$R(\mathbf{?}) \subseteq R(\mathsf{K}) \subset R(\mathsf{MWCh}),$$

the class $R(\mathsf{T})$ can be incorporated in the following way:

$$\boxed{R(\mathbf{?}) \subseteq R(\mathsf{T}) \subseteq R(\mathsf{K}) \subset R(\mathsf{MWCh})}$$

§6. SEQUENCES with COMPLEX STRUCTURE (CHAOTIC SEQUENCES)

The sequences

$$\underline{\underline{(\mathrm{II}_{10})}}:\quad \mathbf{1\,1\,1\,1\ldots1\,1} \qquad \text{и} \qquad \underline{\underline{(\mathrm{III}_{10})}}:\quad \mathbf{1\,0\,1\,0\ldots1\,0},$$

considered above have very simple structure. This simplicity, which results in possibility to describe them easily, justifies the fact that we are inclined to regard these sequences as nonrandom.

On the other hand, one cannot say that the sequence

$$(\mathrm{I}_{10}):\quad \mathbf{0\,1\,1\,1\,0\,1\,0\,0\,1\,0}$$

has simple structure, it is not easy to describe, thus we are inclined to regard it as “random”.

To describe the classes R(MWCh), and R(K), one focuses on the structure of algorithms of creating the $\underline{\underline{\text{sub}}}$sequences.

The **KOLMOGOROV APPROACH**, initiated by him in 1960s, focuses on

the complexity of the structure of SEQUENCES ITSELVES,

either finite or infinite. Kolmogorov introduced certain numerical characteristic (called now the **'Kolmogorov complexity'**) such that

- the complexity of a finite sequence is measured by the length of its shortest "description";
- an infinite sequence is proclaimed to be chaotic (as a synonym of "randomness"), if the complexities of its initial strings grows "as fast as possible".

Now let us turn to formal definitions, including the definition of how one understand a "description".

Each of sequences (I_{10}), (II_{10}), (III_{10}) considered above can be described in words

in DIFFERENT WAYS.

For example, one can say that

- “the word (III_{10}): **1 0 1 0 1 0 1 0 1 0** consists of 10 letters, with 1 at odd positions and 0 at even positions”.

However, it is the same as to say that

- “the word (III_{10}): **1 0 1 0 1 0 1 0 1 0** consists of 10 alternating letters 0 and 1, starting with 1”.

These descriptions have **different lengths**, however it is clear that if one wants to know how to find the shortest then one should

> reduce all these descriptions to a **single standard**
> so that their lengths can be measured in a unified way.

The most natural—and the most simple—way of such a “standardized” description consists in

$$\underline{\textit{coding}\ \ \textit{them}\ \ \textit{in a binary}\ \ \textit{alphabet},}$$

i.e., to represent them as **binary words**.

Thus we shall assume that

> both the words x which we are interested in
> and their coded descriptions y belong to $\boldsymbol{\Xi}$

For a word $y \in \Xi$, its **complexity** is, by definition, the quantity

$$\mathrm{Comp}(y) \equiv \min\{|x| : \ x \text{ is a description of } y\}.$$

It is clear that there may be various “ways of description” (either in prose or in verse. . .). So, it should be determined: what in one or another case is understood under a “way of description”**?**

An approach consists in

considering the so-called **algorithmically computable** mappings $f\colon \Xi \to \Xi$ as “ways of description”.

(We do not cite here all necessary definitions from mathematical logic—see the relevant entries in [Math. Encycl., т. 1] and bibliography therein.)

To a first approximation, the situation may be thought of as follows:

> a “machine”, starting from consequently arriving values $y = (y_1, y_2, \ldots)$, “gives out” certain values $f(y)$ which form a sequence of zeroes and units.

A binary word x is said to be a

“description” (“f-description”) of a finite word y,

if y is the initial fragment of the (finite or infinite) sequence $f(x)$:

$$x = \cdots\cdots \longrightarrow \boxed{\text{MACHINE}} \longrightarrow f(x) = \underbrace{0111001}_{y}\ldots$$

DEFINITION 1. **Complexity** of a word y for a given algorithmically computable mapping f is a number

$$K_f(y) = \text{min}\{|x| \colon x \text{ is } f\text{-description of } y\},$$

where $|x|$ is the length of the binary word x; the minimum of the empty set is assumed to be $+\infty$.

DEFINITION 2. An algorithmically computable mapping f is said to be **optimal**, if for any algorithmically computable mapping g there exists a constant $C = C(g)$ such that

$$K_f(y) = K_g(y) + C \quad \text{for all binary words } y.$$

Kolmogorov and Solomonoff showed that for some important families $\mathcal{F}$ of mappings f such mappings f do exist.

DEFINITION 3. **Entropy** is complexity for an arbitrary optimal algorithmically computable mapping.

Thus, the entropy depends on an optimal mapping. For each optimal mapping there exists its entropy. At first sight this may seem to be an unfavorable circumstance, however any two optimal entropies $K_1(y)$ and $K_2(y)$ differ only by a constant:

$$|K_1(y) - K_2(y)| < C.$$

As we shall see from what follows, to define chaotic sequences it suffices to choose лишь какую-то одну optimal entropy, let us call it $K(y)$.

Note that for the identical mapping $g(y) = y$ the complexity of a word y is evidently equal to its length. Thus, $K(y) \leq |y| + C$ and if $y = (y_1, \dots, y_n)$, then

$$K(y_1, \dots, y_n) \leq n + C.$$

DEFINITION 4. A sequence $y = (y_1, \dots, y_n)$ is called **chaotic** if there exists a constant C such that for any n

$$K(y_1, \dots, y_n) > n - C.$$

This definition shows that the property of a sequence to be chaotic does not depend on a concrete choice of optimal entropy.

In 1973, L. Levin proposed a specification of the class $\mathcal{F}$ of algorithmically computable mappings and studied the corresponding notion of entropy (which is called monotone).

Simultaneously, the same was done by C. P. Schnorr (1973). Their results (taking account of the specifications of the class $\mathcal{F}$) lead to the following important theorem.

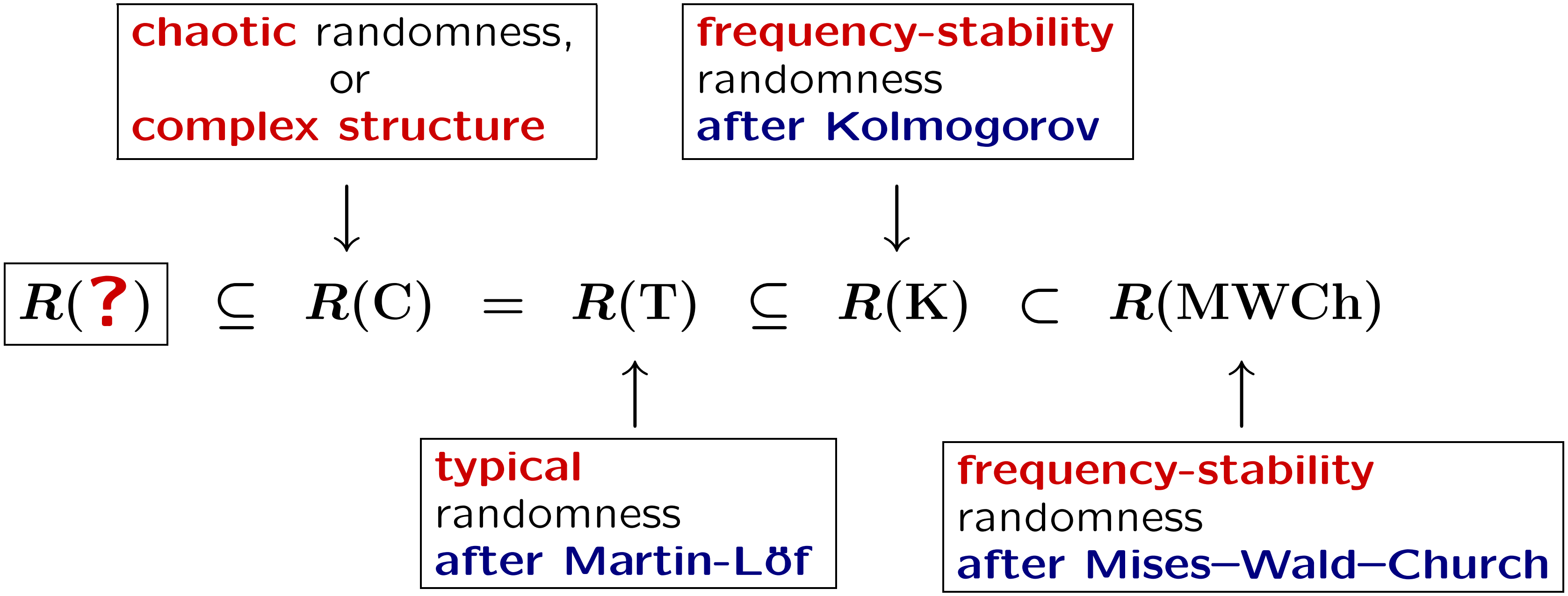
THEOREM (Levin–Schnorr). The class $R(\mathsf{C})$ of chaotic sequences coincides with the class $R(\mathsf{T})$ of typical sequences:

$$\boxed{R(\mathsf{C}) = R(\mathsf{T})}\,.$$

Thus, if we denote

- by $R(\mathrm{T})$ the set of “typical” sequences and
- by $R(\mathrm{C})$ the set of “chaotic” sequences,

then the following diagram hold:



chaotic randomness, or **complex structure**

frequency-stability randomness **after Kolmogorov**

$$\boxed{\boldsymbol{R}(\textbf{?})} \subseteq \boldsymbol{R}(\mathrm{C}) = \boldsymbol{R}(\mathrm{T}) \subseteq \boldsymbol{R}(\mathrm{K}) \subset \boldsymbol{R}(\mathrm{MWCh})$$

typical randomness **after Martin-Lőf**

frequency-stability randomness **after Mises–Wald–Church**

Note that for chaotic-typical random sequences (i.e., sequences which belong to the class $R(\mathrm{C}) = R(\mathrm{T})$) the essential laws of probability theory are valid.

§ 7. NONPREDICTABILITY

As was already noted, Mises did not give an accurate definition of a notion of “random sequences”, and justified their existence by referring to the fact that such sequences are at disposal of casinos. A player which comes to casino, is offered to guess entries of a “random sequence” and gamble on. The casino is convinced of “unpredictability” of such sequences, which does not allow the player to create a strategy which would ruin the gambling house.

J. Ville appears to be the first to use (in a small monograph “Étude critique de la notion de collectif”, Gauthier-Villars, Paris, 1939) the game interpretation to define “unpredictability” as a synonym of “randomness” of a sequence.

Essentially the main objection of Ville to von Mises' notion of collective was the following*:

(a) Given any countable set of place selection function it is possible to construct a sequence $x = (x_1, x_2, \ldots)$ which is a Kollektiv with the property that, for all except finitely many n, $\sum_{k \leq n} x_k \geq n/2$ which is atypical in view of the law of the iterated logarithm (so x is not a sufficiently disorderly sequence).

(b) Von Mises' formalization of gambling strategies (for defending the notion of Kollektiv) as admissible place selection is not perfect, since one may devise a strategy (a **MARTINGALE**) which makes unlimited amounts of many of a sequence of the type constructed in (a), whereas there is no place selection which does this. So, Kollektives are not completely adequate models of random phenomena.

* M. van Lambalgen, **Randomness and foundations of probability: von Mises axiomatization of random sequences**, Probability, statistics and game theory, Institute for Mathematical Statistics, 1996.

Modern views on a “game” approach to the notion of “randomness” via the notion of “unpredictability” can be summarized, after [V.A.Uspensky, The four algorithmic physiognomies of randomness], as follows.

Let a player coming to casino have a certain capital $V(0)$. The casino also possesses a certain capital $W(0)$, whose size is unknown to the player. Let the sequence observed by the player be of the form $x = (x_1, x_2, \ldots)$, where $x_i = \pm 1$. Before the kth step the player choose the size of a stake $\gamma_k = \gamma_k(x_1, \ldots, x_{k-1})$, then his gain/loss at the kth step will be equal to $\gamma_k x_k$, and the total capital will equal

$$V(k) = V(0) + \sum_{i=1}^{k} \gamma_i x_i.$$

Certainly, it is assumed that the stake γ_k at each time k cannot exceed the capital $V(k-1)$, i.e., γ_k is subject to the restriction $\gamma_k \le V(k-1)$. The player has a right to take $\gamma_k = 0$, which means that he does not stake at all. In this case the player capital remains unchanged.

By definition, the player wins, if

$$\sup_k V(k) = \infty,$$

i.e., without regard to the capital W in possession of casino, there comes a time when casino finds itself ruined.

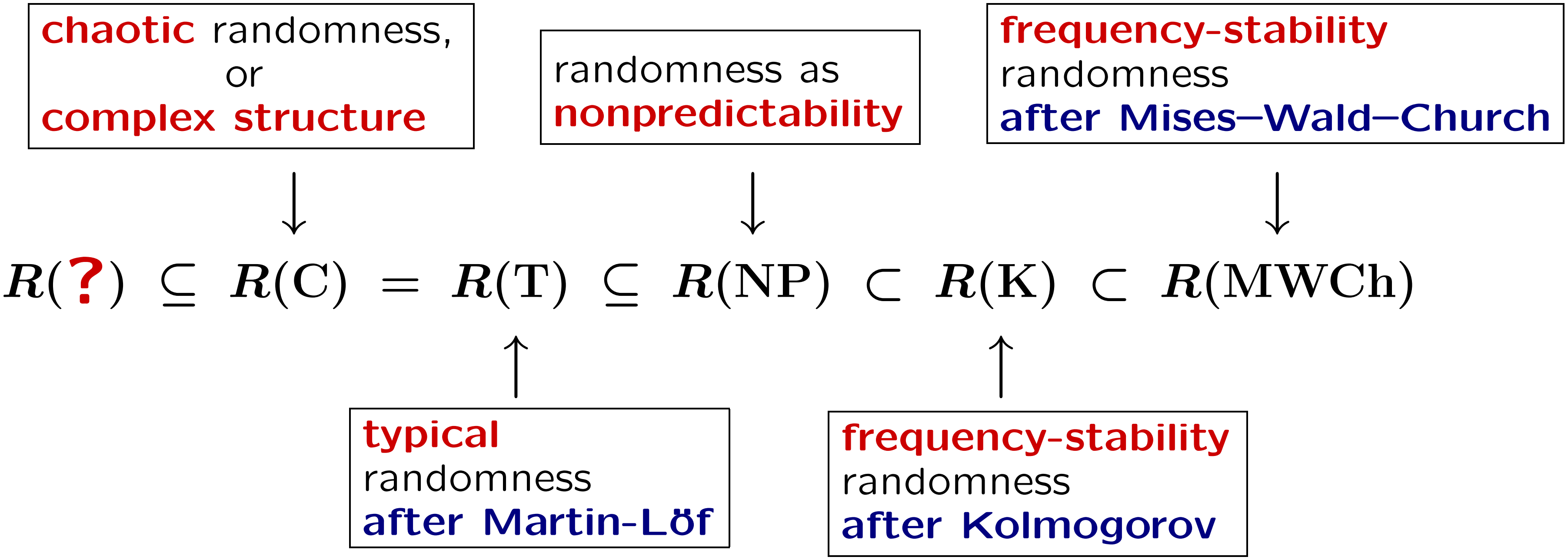
It is important to emphasize that all the strategies (γ_k) of the player are supposed computable, i.e. are given by means of a certain algorithm.

The sequence $x = (x_1, x_2, \ldots)$ is called **predictable**, if there exists a winning computable strategy (γ_k), otherwise the sequence is called **nonpredictable**. Denote by $R(\mathsf{NP})$ the class of nonpredictable sequences. It is known that

$$R(\mathsf{T}) \subseteq R(\mathsf{NP}) \subset R(\mathsf{K})$$

(note that the second inclusion is proper (!)).

So, we can sum up the state of our knowledge about relations between different classes $R(\cdot)$ of sequences:



chaotic randomness, or **complex structure**

randomness as **nonpredictability**

frequency-stability randomness **after Mises–Wald–Church**

$$\boldsymbol{R}(\textbf{?}) \subseteq \boldsymbol{R}(\mathrm{C}) = \boldsymbol{R}(\mathrm{T}) \subseteq \boldsymbol{R}(\mathrm{NP}) \subset \boldsymbol{R}(\mathrm{K}) \subset \boldsymbol{R}(\mathrm{MWCh})$$

typical randomness **after Martin-Löf**

frequency-stability randomness **after Kolmogorov**

Note that it is so far unknown whether the equality $\underline{\underline{\boldsymbol{R}(\mathrm{T}) \overset{?}{=} \boldsymbol{R}(\mathrm{NP})}}$ holds. This problem is still waiting to be solved.

§ 8. AN EXAMPLE

We want to demonstrate now

how the notions of **algorithmic theory of probabilities** allows one to give new proofs of some results of the **classical theory of probabilities**.

As an example, we take

the STRONG LAW OF LARGE NUMBERS:

$$\frac{S_n(x)}{n} \to \frac{1}{2}, \qquad \text{where} \quad x = (x_1, x_2, \ldots), \quad x_i = 0 \text{ or } 1$$

The ideas of the algorithmic proof (**V. Vovk, A. Shen**) are the following.

Take the initial segment $x_{(n)} = (x_1, \ldots, x_n)$, and let $p_n = \frac{1}{n}(x_1 + \cdots + x_n)$ be a frequency of units in this segment.

From Shannon we know that "**entropy per letter**" in $x_{(n)}$ [i.e., quantity of bits necessary for encoding one letter in $x_{(n)}$] is about $\boxed{H(p_n) \equiv -p_n \log_2 p_n - q_n \log_2 q_n}$.
So, we need $nH(p_n)$ bits to encode $x_{(n)}$.

However, to encode or decode $x_{(n)}$, we must know also probability p_n. Hence the full code for $x_{(n)}$ includes also the number p_n, which is a rational number whose numerator and denominator do not exceed n. So, its encoding requires not more than $O(\log n)$ bits.

Therefore, the initial segment $x_{(n)}$ of any sequence $(x_0, x_1, \ldots)$ contains not more than

$$nH(p_n) + O(\log n) \quad \text{bits of information.}$$

If the sequence $(x_0, x_1, \ldots)$ is **chaotic**, then, by Kolmogorov–Levin, the monotone entropy (complexity) of its segment of the length n must be $n + O(1)$. So,

$$H(p_n) = \frac{n + O(1) + O(\log n)}{n} = 1 + O\left(\frac{\log n}{n}\right) \longrightarrow 1, \qquad n \to \infty,$$

and therefore, $p_n \to 1/2$ (because $H(p) = 1 + \mathrm{const}(p - 1/2)^2 + o((p - 1/2)^2)$ near $p = 1/2$, so that $p_n - 1/2 = O(\sqrt{n^{-1} \log n})$).

Summary: for all chaotic (=typical) sequences $(x_0, x_1, \ldots)$

$$p_n = \frac{x_1 + \cdots + x_n}{n} \longrightarrow \frac{1}{2}.$$

These sequences form a set of measure 1, so the **classical** strong law of large numbers is proved. $\square$