40 YEARS OF THE ORIGIN OF KOLMOGOROV RANDOMNESS THEORY

Alexei Semenov Andrej Muchnik

Institute of New Technologies
Moscow

A. N. Kolmogorov. On tables of random numbers.
Sankhyã: The Indian Journal of Statistics, Series A, 1963, v. 25, part 4.
(Reprinted in *Theoretical Computer Science*, 1998, v. 207, pp. 387–395).

"The set theoretic axioms of the calculus of probability . . . had solved the majority of formal difficulties . . . so successfully that the problem of finding the basis of real applications . . . became rather secondary to many investigators."

R. von Mises	frequency approach to randomness (1919)	infinite
A. Church	computable monotone rules (1940)	sequences
A. Kolmogorov	combinatorial (1963) algorithmical entropy (1965)	finite sequences

Probability of 0 $\quad p = \frac{1}{2}$

Binary sequences of length L
rule r:

$$\mathbf{t} = \overbrace{\boxed{\mathsf{X}}\ \_\ \boxed{\mathsf{X}}\ \dots \boxed{0}\ \boxed{\mathsf{X}}\ \boxed{1}\ \boxed{0}\ \boxed{\mathsf{X}}\ \_\ \boxed{\mathsf{X}}\ \boxed{1}\ \dots \boxed{\mathsf{X}}\ \_\ \boxed{0}}^{L}$$

$$r[\mathbf{t}] = \underbrace{\boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{0}\ \dots \boxed{1}\ \boxed{1}}_{n}$$

R. von Mises	frequency approach to randomness (1919)	infinite
A. Church	computable monotone rules (1940)	sequences
A. Kolmogorov	combinatorial (1963) algorithmical entropy (1965)	finite sequences

Probability of 0 $\quad p = \frac{1}{2}$

Binary sequences of length L
rule r:

$$\mathbf{t} = \overbrace{\boxed{X}\ \_\ \boxed{X} \dots \boxed{0}\ \boxed{X}\ \boxed{1}\ \boxed{0}\ \boxed{X}\ \_\ \boxed{X}\ \boxed{1} \dots \boxed{X}\ \_\ \boxed{0}}^{L}$$

$$r[\mathbf{t}] = \underbrace{\boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{0} \dots \boxed{1}\ \boxed{1}}_{n}$$

R. von Mises	frequency approach to randomness (1919)	infinite
A. Church	computable monotone rules (1940)	sequences
A. Kolmogorov	combinatorial (1963) algorithmical entropy (1965)	finite sequences

Probability of 0 $\quad p = \frac{1}{2}$

Binary sequences of length L

rule r:

$$\mathbf{t} = \overbrace{\boxed{\mathrm{X}}\ \_\ \boxed{\mathrm{X}} \ldots \boxed{0}\ \boxed{\mathrm{X}}\ \boxed{1}\ \boxed{0}\ \boxed{\mathrm{X}}\ \_\ \underline{1}\ \boxed{1} \ldots \boxed{\mathrm{X}}\ \_\ \boxed{0}}^{L}$$

$$r[\mathbf{t}] = \underbrace{\boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{0} \ldots \boxed{1}\ \boxed{1}\ \boxed{1}}_{n+1}$$

Monotone rules open cards in the original order.

Nonadaptive rules do not use information about digits on the previously opened cards.

Main definition:

$\mathfrak{R}_L$ is a system of rules

sequence $\mathbf{t}$ is (n, ε)-random w.r.t. $\mathfrak{R}_L$ if
for any $r \in \mathfrak{R}_L$

$$\ell(r[\mathbf{t}]) \geqslant n \quad \Rightarrow \quad \left| \mathrm{freq}_0(r[\mathbf{t}]) - \frac{1}{2} \right| < \varepsilon$$

(random sequence = table of random numbers = random generator) number of (n, ε)-nonrandom sequences w.r.t. one rule

$$\approx 2^{L} \mathrm{e}^{-2n\varepsilon^2}$$

Abbreviation:

$$d(n, \varepsilon) = 2n\varepsilon^2 \log_2 \mathrm{e}$$

1. "The frequency concept based on the notion of limiting frequency as the number of trials increases to infinity, does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials."

2. "The frequency concept applied to a large but finite number of trials does not admit a rigorous formal exposition within the framework of pure mathematics."

(1) "I still maintained the first of the two theses mentioned above."

(2) "Such a conception in its full development requires the introduction of a measure of the complexity of the algorithm. I shall use the fact that there cannot be a *very large number of simple algorithms*."

Kolmogorov (1965):

Three approaches to the definition of the concept "quantity of information".

- The approach based on Kolmogorov axiomatization of probability theory (studied by Shannon since 1940-s).
- The combinatorial approach.
- The algorithmical approach.

The comparision of the combinatorial and algorithmical approaches is based on the following statements.

Proposition 1 (Kolmogorov). The number of objects with conditional entropy w.r.t. a fixed object smaller than m is not greater than $2^m - 1$.

Proposition 2 (Kolmogorov). If a set of cardinality 2^m can be enumerated by a program with entropy smaller than α, then entropy of an arbitrary element of the set is smaller than $m + \alpha + 2\log_2 \alpha + C$.

(Constant C depends only on choice of optimal programming language.)

(Entropy can be conditional w.r.t. a fixed object.)

Combinatorial complexity of a finite set is binary logarithm of its cardinality.

Kolmogorov (arbitrary p):

$l(n,\varepsilon)$ is the maximal number l:
$\forall\, p$, $\forall\, L$, $\forall$ set of rules of combinatorial complexity $< l$,
$\exists$ table of (n,ε,p)-random numbers of length L

Kolmogorov:
$\forall\, \varepsilon$ small enough (e. g., $\varepsilon < 1/20$) $\forall\, n \geqslant \varepsilon^{-4}$

$$(2\log_2 \mathrm{e})n\varepsilon^2(1-\varepsilon) < l(n,\varepsilon) < 4n\varepsilon(1+5\varepsilon) \qquad (*)$$

Kolmogorov:
the problem of removing the discrepancy in the power of ε.

Theorem.

$$l(n,\varepsilon) < (2\log_2 \mathrm{e})n\varepsilon^2(1+2\varepsilon)$$

for ε small enough and $n \geqslant \varepsilon^{-4}$.

back to $p = \frac{1}{2}$

Theorem 1 (Kolmogorov).
For every system of rules $\mathfrak{R}_L$:

$$\text{complexity of } \mathfrak{R}_L < d(n,\varepsilon)(1-\varepsilon)$$
$$\Downarrow$$
$$\exists\ (n,\varepsilon)\text{-random sequence}$$

(probabilistic proof)

Theorem 2 (Kolmogorov).
$\exists$ system $\mathfrak{R}_L$

$$\text{complexity of } \mathfrak{R}_L < 4n\varepsilon(1+5\varepsilon),$$
$$\nexists\ (n,\varepsilon)\text{-random sequence w.r.t. } \mathfrak{R}_L$$

(explicit construction)

Kolmogorov:

"Unfortunately, I cannot remove the discrepancy between the power of ε in $2n\varepsilon^2 + o(n\varepsilon^2)$ and $4n\varepsilon + o(n\varepsilon)$."

Theorem 1 (Kolmogorov).
For every system of rules $\mathfrak{R}_L$:

$$\text{complexity of } \mathfrak{R}_L < d(n,\varepsilon)(1-\varepsilon)$$
$$\Downarrow$$
$$\exists\ (n,\varepsilon)\text{-random sequence}$$

(probabilistic proof)

Theorem 3.
$\exists$ system $\mathfrak{R}_L$

$$\text{complexity of } \mathfrak{R}_L < d(n,\varepsilon)\frac{1+\varepsilon}{1-n/(L-1)},$$
$$\nexists\ (n,\varepsilon)\text{-random sequence w.r.t. } \mathfrak{R}_L$$

(probabilistic proof)

$$\varepsilon = \frac{1}{5} \qquad\qquad n = 10000 \qquad\qquad L = 100000$$

Combinatorial complexity of a test $\leqslant 923 \;\Rightarrow\;$ random sequence exists

For some test of combinatorial complexity $\geqslant 1539$

random sequence does not exist

combinatorial vs. algorithmical

Theorem 1 (Kolmogorov).
For every system of rules $\mathfrak{R}_L$:
complexity of $\mathfrak{R}_L < d(n,\varepsilon)(1-\varepsilon)$

$$\Downarrow$$

$\exists$ (n,ε)-random sequence.

(probabilistic proof)

Theorem 1′.
For the system $\mathfrak{R}_L$ consisting of all rules having conditional entropy with respect to L

$$< d(n,\varepsilon)(1-\varepsilon)\,,$$

$\exists$ (n,ε)-random sequence.

Theorem 1 (Kolmogorov).
For every system of rules $\mathfrak{R}_L$:
complexity of $\mathfrak{R}_L < d(n,\varepsilon)(1-\varepsilon)$

$$\Downarrow$$

$\exists$ (n,ε)-random sequence.

Theorem 1′.
For the system $\mathfrak{R}_L$ consisting of all rules having conditional entropy with respect to L

$$< d(n,\varepsilon)(1-\varepsilon)\,,$$

$\exists$ (n,ε)-random sequence.

Theorem 3.
$\exists$ system $\mathfrak{R}_L$ of complexity

$$< d(n,\varepsilon)\frac{1+\varepsilon}{1-n/(L-1)}$$

such that no sequence is (n,ε)-random w.r.t. $\mathfrak{R}_L$.

Theorem 3′.
For the system $\mathfrak{R}_L$ of all non-adaptive rules having conditional entropy with respect to L

$$< d(n,\varepsilon)\frac{1+\varepsilon}{1-n/(L-1)}+C\,,$$

there is no (n,ε)-random sequence.

For any sequence $\mathfrak{t}$ of length L

the frequencies of zeros and ones in all long enough subsequences selected from $\mathfrak{t}$ by simple enough rules are close to $1/2$

$$\Downarrow$$

the entropy of $\mathfrak{t}$ conditional to L is greater than $L(1 - \ln\ln L/\ln L)$.

The reverse implication is easy to prove,
this was noted by Kolmogorov (1982)