



Soviet-era science, translated into English

Reports of the Academy of Sciences of the USSR

Academician A. N. KOLMOGOROV

1958

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-195801.96025>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

Reports of the Academy of Sciences of the USSR
1958. Volume 119, No. 5

MATHEMATICS

Academician A. N. KOLMOGOROV

A NEW METRIC INVARIANT OF TRANSITIVE DYNAMICAL SYSTEMS AND AUTOMORPHISMS OF LEBESGUE SPACES

It is well known that a substantial part of the metric theory of dynamical systems can be presented as an abstract theory of “flows” $\{S_t\}$ on “Lebesgue spaces” M with measure μ , in terms invariant with respect to “isomorphisms modulo zero” (see the survey article by V. A. Rokhlin ⁽¹⁾, to which the following exposition adheres as regards definitions and notation). We shall assume the measure on M to be normalized by the condition

$$\mu(M) = 1 \tag{1}$$

and nontrivial (i.e., assume the existence of a set $A \subseteq M$ with $0 < \mu(A) < 1$). Many examples are known of transitive automorphisms and transitive flows with the so-called “countable Lebesgue spectrum” (for automorphisms see ⁽¹⁾, § 4; for flows, ⁽²⁻⁵⁾). From the spectral point of view we have here one type of automorphisms Ω_0^ω and one type of flows Ω^ω . The question whether all automorphisms of type Ω_0^ω (respectively, flows of type Ω^ω) are isomorphic to one another mod 0 has until now remained open. We show in §§ 3-4 that the answer to this question is negative both in the case of automorphisms and in the case of flows. The new invariant that makes it possible to split the class of automorphisms Ω_0^ω and the class of flows Ω^ω into a continuum of invariant subclasses is **entropy per unit time**. In § 1 the necessary facts from information theory are presented (the notions of conditional entropy and conditional information introduced here and their properties are probably of broader interest as well, although the entire exposition directly adjoins the definition of the amount of information from ⁽⁷⁾ and the numerous works developing that definition). In § 2 the characteristic h is defined and its invariance is proved. In §§ 3-4 examples are indicated of automorphisms and flows with arbitrary values of h in the range $0 < h \leq \infty$. In the case of automorphisms these are long-constructed examples; in the case of flows the construction of examples with finite h is a more delicate problem connected with certain curious questions in the theory of Markov processes.

§ 1. Properties of conditional entropy and conditional amount of information. In accordance with (1), denote by \mathfrak{S} the Boolean algebra of measurable sets of the space M , considered mod 0. Let \mathfrak{C} be a subalgebra of the algebra \mathfrak{S} closed in the metric $\rho(A, B) = \mu((A - B) \cup (B - A))$. It generates a partition $\xi_{\mathfrak{C}}$ of the space M , defined mod 0, determined by the condition that $A \in \mathfrak{C}$ if and only if mod 0 all of A can be composed of entire elements of the partition $\xi_{\mathfrak{C}}$. On the elements C of the partition $\xi_{\mathfrak{C}}$ the “canonical system of measures μ_C ”⁽¹⁾ is defined. For any $x \in C$ we shall put

$$\mu_x(A | \mathfrak{C}) = \mu_C(A \cap C). \quad (2)$$

From the point of view of probability theory (where any measurable function of an element $x \in M$ is called a “random variable”), the random variable $\mu_x(A | \mathfrak{C})$ is the “conditional probability” of the event A given the known outcome of the “experiment” \mathfrak{C} ⁽⁶⁾, Ch. I, § 7).

For three subalgebras \mathfrak{A} , \mathfrak{B} , and \mathfrak{C} of the algebra \mathfrak{S} , and for $C \in \xi_{\mathfrak{C}}$, set

$$I_C(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) = \sup \sum_{i,j} \mu_x(A_i \cap B_j) \log \frac{\mu_x(A_i \cap B_j)}{\mu_x(A_i)\mu_x(B_j)}, \quad (3)$$

where the least upper bound is taken over all finite decompositions $M = A_1 \cup A_2 \cup \dots \cup A_n$, $M = B_1 \cup B_2 \cup \dots \cup B_n$, for which $A_i \cap A_j = N$, $B_i \cap B_j = N$, $i \neq j$, $A_i \in \mathfrak{A}$, $B_j \in \mathfrak{B}$ (N is the empty set). If \mathfrak{C} is the trivial algebra $\mathfrak{N} = \{N, M\}$, then (3) becomes the definition of the unconditional information $I(\mathfrak{A}, \mathfrak{B})$ from Appendix 7 to (7)*. The quantity (3) itself is interpreted as the “amount of information in the results of the experiment \mathfrak{A} relative to the experiment \mathfrak{B} , given the known outcome C of the experiment \mathfrak{C} .” If one does not fix $C \in \xi_{\mathfrak{C}}$, then it is natural to consider the random variable $I(\mathfrak{A}, \mathfrak{B} | \mathfrak{C})$, which for $x \in C$ is equal to $I_x(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) = I_C(\mathfrak{A}, \mathfrak{B} | \mathfrak{C})$. In what follows we shall deal with its mathematical expectation

$$\mathbf{M}I(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) = \int_M I_x(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) \mu(dx). \quad (4)$$

The definitions of conditional entropy and of mean conditional entropy require no special explanation: $H(\mathfrak{A} | \mathfrak{C}) = I(\mathfrak{A}, \mathfrak{A} | \mathfrak{C})$, $\mathbf{M}H(\mathfrak{A}, \mathfrak{C}) = \int_M H_x(\mathfrak{A} | \mathfrak{C}) \mu(dx)$.

We note those properties of the conditional amount of information and conditional entropy which we shall need below. Properties (α) and (δ) for the case of unconditional amount of information and entropy are well known; property (ε) for the unconditional amount of information constitutes the content of Theorem 2 of note (8). Properties (β) and (γ) are proved without difficulty. Concerning property (β) , it should only be noted that the analogous assertion for the

amount of information (from $\mathfrak{C} \supseteq \mathfrak{C}'$ there follows: $I(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) \leq I(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}')$) would already be erroneous. Connected with this is the fact that in property (ζ) there stands the lower limit and the sign \geq : the corresponding limit may fail to exist, and the lower limit may in some cases turn out to be greater than $MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C})$.

(α) $I(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) \leq H(\mathfrak{A} | \mathfrak{C})$, and equality is certainly attained when $\mathfrak{B} \supseteq \mathfrak{A}$.

(β) If $\mathfrak{C} \supseteq \mathfrak{C}'$, then $H(\mathfrak{A} | \mathfrak{C}) \leq H(\mathfrak{A} | \mathfrak{C}')$, mod 0.

(γ) If $\mathfrak{B} \supseteq \mathfrak{B}'$, then

$$MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) = MI(\mathfrak{A}, \mathfrak{B}' | \mathfrak{C}) + MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C} \vee \mathfrak{B}'),$$

where $\mathfrak{C} \vee \mathfrak{B}'$ is the minimal closed algebra containing \mathfrak{C} and \mathfrak{B}' .

(δ) If $\mathfrak{B} \supseteq \mathfrak{B}'$, then

$$MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}) \geq MI(\mathfrak{A}, \mathfrak{B}' | \mathfrak{C}).$$

(ε) If

$$\mathfrak{A}_1 \subseteq \dots \subseteq \mathfrak{A}_n \subseteq \dots, \quad \bigcup \mathfrak{A}_n = \mathfrak{A},$$

then

$$\lim_{n \rightarrow \infty} MI(\mathfrak{A}_n, \mathfrak{B} | \mathfrak{C}) = MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}).$$

(ζ) If

$$\mathfrak{C}_1 \supseteq \mathfrak{C}_2 \supseteq \dots \supseteq \mathfrak{C}_n \supseteq \dots, \quad \bigcap_n \mathfrak{C}_n = \mathfrak{C},$$

then

$$\liminf_{n \rightarrow \infty} MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}_n) \geq MI(\mathfrak{A}, \mathfrak{B} | \mathfrak{C}).$$

§ 2. Definition of the invariant h . We shall say that a flow $\{S_t\}$ is quasiregular (has type \mathfrak{R}) if** there exists a closed subalgebra \mathfrak{S}_0 of the algebra \mathfrak{S} , whose shifts $\mathfrak{S}_t = S_t \mathfrak{S}_0$ have the following properties: (I) $\mathfrak{S}_t \subseteq \mathfrak{S}_{t'}$, if $t \leq t'$. (II) $\bigcup_t \mathfrak{S}_t = \mathfrak{S}$. (III) $\bigcap_t \mathfrak{S}_t = \mathfrak{R}$.

* The authors of note ⁽⁸⁾ did not in due time pay attention to Appendix 7 to ⁽⁷⁾, included in the Russian translation ⁽⁹⁾. Note ⁽⁸⁾ should have begun with a reference to this appendix to ⁽⁷⁾.

** This condition is considerably weaker than the condition of “regularity” usually used in the theory of random processes. See about this at the end of § 4.

When the flow is interpreted as a stationary random process, \mathfrak{S}_t may be regarded as the algebra of events “depending only on the course of the process up to the moment of time t .” It is easily proved that flows of type \mathfrak{R} are transitive, and from Plessner’s results ^(10, 11) one can infer that they have homogeneous Lebesgue spectrum. If the multiplicity of the spectrum is equal to ν ($\nu = 1, 2, \dots, \omega$), then

we assign the flow to the type \mathfrak{R}^ν . Clearly, $\mathfrak{R}^\nu \subset \Omega^\nu$, where Ω^ν is the class of flows with Lebesgue spectrum of homogeneous multiplicity ν . It is possible, however, that all Ω^ν (and, consequently, \mathfrak{R}^ν), except for $\Omega^\omega(\mathfrak{R}_0^\omega)$, are empty and that $\mathfrak{R}^\omega = \Omega^\omega$.

Theorem 1. *If for the flow $\{S_t\}$ there exists an \mathfrak{S}_0 satisfying conditions (I)–(III), then for $\Delta > 0$*

$$\mathbf{MH}(\mathfrak{S}_{t+\Delta} \mid \mathfrak{S}_t) = h\Delta,$$

where h is a constant lying in the range $0 < h \leq \infty$.

Theorem 2. *The constant h for a given flow $\{S_t\}$ does not depend on the choice of \mathfrak{S}_0 satisfying conditions (I)–(III).*

We outline here the proof of Theorem 2. Let two \mathfrak{S}_0 and \mathfrak{S}'_0 correspond to $h < \infty$ and h' . By Theorem 1 and lemmas (α) and (ε) , for every $\varepsilon > 0$ one can find such a k that

$$h = \mathbf{MH}(\mathfrak{S}_{t+1} \mid \mathfrak{S}_t) = \mathbf{MI}(\mathfrak{S}_{t+1}, \mathfrak{S} \mid \mathfrak{S}_t) \leq \mathbf{MI}(\mathfrak{S}_{t+1}, \mathfrak{S}'_{t+k} \mid \mathfrak{S}_t) + \varepsilon. \quad (5)$$

From (5), by lemma (ζ) , it follows that there exists such an m that

$$h \leq \mathbf{MI}(\mathfrak{S}_{t+1}, \mathfrak{S}'_{t+k} \mid \mathfrak{S}_t \vee \mathfrak{S}'_s) + 2\varepsilon \quad \text{for } t - s > m. \quad (6)$$

From (6) and lemmas (δ) , (γ) , (α) , (β) (to be applied in the indicated order!):

$$\begin{aligned} nh &\leq \sum_{t=0}^{n-1} \mathbf{MI}(\mathfrak{S}_{t+1}, \mathfrak{S}'_{t+k} \mid \mathfrak{S}_t \vee \mathfrak{S}'_{-m}) + 2n\varepsilon \leq \\ &\leq \sum_{t=0}^{n-1} \mathbf{MI}(\mathfrak{S}_{t+1}, \mathfrak{S}'_{n+k} \mid \mathfrak{S}_t \vee \mathfrak{S}'_{-m}) + 2n\varepsilon = \\ &= \mathbf{MI}(\mathfrak{S}_x, \mathfrak{S}'_{n+k} \mid \mathfrak{S}_0 \vee \mathfrak{S}'_{-m}) + 2n\varepsilon \leq \\ &\leq \mathbf{MH}(\mathfrak{S}'_{n+k} \mid \mathfrak{S}_0 \vee \mathfrak{S}'_{-m}) + 2n\varepsilon \leq \\ &\leq \mathbf{MH}(\mathfrak{S}'_{n+k} \mid \mathfrak{S}'_{-m}) + 2n\varepsilon = \\ &= (n + k + m)h' + 2n\varepsilon, \\ h &\leq \frac{n + k + m}{n} h' + 2\varepsilon. \end{aligned} \quad (7)$$

Since $\varepsilon > 0$ and n are arbitrary (with n chosen after k and m have been fixed), (7) implies the inequality $h \leq h'$. This inequality is proved quite analogously also in the case $h = \infty$. The reverse inequality $h' \leq h$ is proved similarly, and this completes the proof of Theorem 2.

§ 3. Invariants of automorphisms. If in § 2 we assume that t takes only integer values, then $\{S_t\}$ is uniquely determined by the automorphism $T = S_1$. By Theorems 1 and 2 there exists an invariant $0 < h(T) \leq \infty$.

It is easily proved that every automorphism of type \mathfrak{R}_0 (the subscript is used to distinguish it from the case of flows with continuous time) has countably multiple Lebesgue spectrum, i.e. among the classes \mathfrak{R}'_0 only the class $\mathfrak{R}^\omega_0 \subset \Omega^\omega_0$ is nonempty. It decomposes according to the values of $h(T)$ into classes $\mathfrak{R}^\omega_0(h)$.

Theorem 3. *For every h , $0 < h \leq \infty$, there exists an automorphism belonging to $\mathfrak{R}^\omega_0(h)$.*

The corresponding examples are well known and are obtained, for example, from the scheme of independent random trials $\Omega_{-1}, \Omega_0, \Omega_1, \dots, \Omega_t, \dots$ with distribution of the probabilities of the outcome ξ_t of the trial Ω_t

$$\mathbf{P}\{\xi_t = a_i\} = p_i, \quad - \sum_{i=1}^{\infty} p_i \log p_i = h. \quad (8)$$

The space M consists of sequences $x = (\dots, x_{-1}, x_0, x_1, \dots, x_t, \dots)$, $x_t = a_1, a_2, \dots$, and the shift $Tx = x'$ is defined by the formula $x'_t = x_{t-1}$. The measure μ on M is defined as the direct product of the probability measures (8).

§ 4. Invariants of flows.

Theorem 4. *For any h , $0 < h < \infty$, there exists a flow of class $\mathfrak{R}^\omega(h)$, i.e. a flow with countably multiple Lebesgue spectrum and with the prescribed value of the constant h .*

By analogy with § 3 there naturally arises the idea of using, for the proof of Theorem 4, instead of the scheme of discrete independent trials, the scheme of “processes with independent increments,” or of generalized processes “with independent values”^(12,13). However, this path leads only to flows of the class $\mathfrak{R}^\omega(\infty)$ ⁽⁵⁾. To obtain finite values of h one has to use a more artificial construction. In this note it is possible only to give a **description** of one such construction.

Let us define mutually independent random variables ξ_n , corresponding to all integers n , with distributions of their values: $\mathbf{P}(\xi_0 = k) = 3/4^k$, $k = 1, 2, \dots$, and, for $n \neq 0$, $\mathbf{P}\{\xi_n = k\} = 1/2^k$, $k = 1, 2, \dots$. We place the point τ_0 on the t -axis, in the case $\xi_0 = k$, with uniform distribution of probabilities on the interval $-u/2^k \leq \tau_0 \leq 0$, and define the points τ_n for $n \neq 0$ from the relation

$$\tau_{n+1} = \tau_n + u/2^{\xi_n}.$$

Put $\varphi(t) = \xi_n$ for $\tau_n \leq t < \tau_{n+1}$. It is easy to verify that the distribution of the random function $\varphi(t)$ is invariant with respect to the shifts $S_t\varphi(t_0) = \varphi(t_0 - t)$. It is easy to calculate that $h\{S_t\} = 6/u$ (per unit of time there fall on average $3/u$ points τ_n , and each ξ_n contributes entropy

$$\sum_{k=1}^{\infty} \frac{k}{2^k} = 2$$

).

One can obtain a more visual representation of our random process if one includes in the description of its state $\omega(t)$ at the moment of time t , in addition to the value $\varphi(t)$, also the value $\delta(t) = t - \tau^*(t)$ of the difference between t and the nearest point τ_n to the left of t . With this method of description our process turns out to be a stationary **Markov** process. It deserves only the name “quasiregular,” since, although the corresponding dynamical system is transitive, the value of the difference $f(\omega(t), t) = \tau^*(t) = t - \delta(t)$ is determined with the precision of a dyadic-rational term by the behavior of a realization of the process in the arbitrarily distant past.

Received
21 I 1958

REFERENCES

- ¹ V. A. Rokhlin, *Uspekhi Mat. Nauk*, **4**, 2 (30) (1949).
- ² I. M. Gel' fand, S. V. Fomin, *Uspekhi Mat. Nauk*, **7**, 1 (47) (1952).
- ³ S. V. Fomin, *Ukr. Mat. Zhurn.*, **2**, No. 2 (1950).
- ⁴ K. Itô, *Japan J. Math.*, **22**, 63 (1952).
- ⁵ K. Itô, *Trans. Am. Math. Soc.*, **81**, 253 (1956).
- ⁶ J. L. Doob, *Stochastic Processes*, IL, 1956.
- ⁷ C. E. Shannon, W. Weaver, *The Mathematical Theory of Communications*, 1949.
- ⁸ I. M. Gel' fand, A. N. Kolmogorov, A. M. Yaglom, *DAN*, **111**, No. 4 (1956).
- ⁹ K. Shannon, Collected papers, *Theory of Transmission of Electrical Signals in the Presence of Noise*, IL, 1953.
- ¹⁰ A. I. Plesner, *DAN*, **23**, No. 4 (1939).
- ¹¹ A. I. Plesner, *DAN*, **25**, No. 9 (1939).
- ¹² K. Itô, *Mem. Coll. Sci. Univ. Kyoto*, **18**, No. 3 (1954).
- ¹³ I. M. Gel' fand, *DAN*, **100**, No. 5 (1955).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.