A. N. Kolmogorov's and Y. G. Sinai's papers introducing entropy of dynamical systems

Contents

Translator's Preface

The discovery of the notion of entropy of dynamical systems is quite interesting and instructive, and has some unexpected twists. The story started with the Kolmogorov's idea of transplanting Shannon's notion of entropy from the theory of information into the theory of dynamical systems, and with Kolmogorov's breakthrough results about Bernoulli shifts. For the first publication of his definition of entropy and its applications Kolmogorov developed a new approach intended to work not only for the dynamical systems with discrete time (i.e. automorphisms) such as Bernoulli shifts, but also for the dynamical systems with continuous time (i.e. flows). This approach was based on a theorem which turned out to be wrong, as it was very soon pointed out by Rokhlin. Both Kolmogorov and Sinai (who was at the time a doctoral student of Kolmogorov) quickly and independently repaired the theory for automorphisms. Surprisingly, the case of the flows turned out to be more difficult, but nevertheless within two months Sinai introduced a working definition of the entropy of flows.

Sinai published in English at least two accounts of these events; see [1], [2]. But the original 1958 and 1959 papers by Kolmogorov and Sinai were never translated. The only exception is the first 1959 paper of Sinai, which was only recently translated by Sinai himself for the inclusion into the first volume of his *Selecta*; see [3]. Clearly, all these papers are important historical documents and deserve to be translated into English. This is done in the present collection of translations. The already translated 1959 paper of Sinai is translated anew, but only partially. Namely, only the part devoted to the definition of entropy is translated. In contrast with the free translation by Sinai, the present translation is intended to be as faithful to the original as possible. In addition, 1985 version of Kolmogorov's 1958 paper and his 1987 comments to the same paper were translated. The 1985 version of Kolmogorov's paper was translated anew without consulting the existing translation. The translator hopes the present translation is more faithfull to the original than the previous one.

The translator attempted to preserve to the extent possible even the formatting features of Russians originals. Some of them are quite outdated, as is the writing style of *Doklady* 1958-1959 papers, which the translator also attempted to follow.

References

[1] Y. G. Sinai, Kolmogorov's work on ergodic theory, *The Annals of Probability*, Vol. 17 (1989), No. 3, 833–839.

[2] Y. G. Sinai, Metric entropy of dynamical systems, Preprint, 2007, http://web.math.princeton.edu/facultypapers/Sinai/.

[3] Y. G. Sinai, On the notion of entropy of a dynamical system, A free translation from the Russian original by the author, *Selecta, Volume I: Ergodic Theory and Dynamical Systems*, Springer Science+Business Media, LLC, 2010, 3–8.

Translator: http://nikolaivivanov.com

A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces*†

A. N. Kolmogorov

It is well known that a significant part of the metric theory of dynamical systems can be presented as an abstract theory of "flows" $\{S_t\}$ on "Lebesgue spaces" M with a measure μ in a form, invariant with respect to "isomorphisms modulo zero" (see the expository paper [1] by V. A. Rokhlin, the definitions and notations of which are adopted in the following exposition). We will assume that the measure on M is normalized by the condition

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

and is non-trivial (i.e., we assume that there exist a set $A \subset M$ with $0 < \mu(A) < 1$). Many examples of transitive automorphisms and transitive flows having so-called Lebesgue spectrum of countably infinite multiplicity are known (for automorphisms see [1, § 4], for flows [2 – 5]). From the spectral point of view we have here only one type of automorphisms $\mathcal{L}_0^\omega$ and only one type of flows $\mathcal{L}^\omega$. The question if all automorphisms of type $\mathcal{L}_0^\omega$ (respectively, all flows of type $\mathcal{L}^\omega$) are metrically isomorphic to each other mod 0 remained open up to now. We show in § 3, § 4 that the answer to this question negative both in the case of automorphisms and in the case of flows. The new invariant, which allows to split the class of automorphisms $\mathcal{L}_0^\omega$ and the class of flows $\mathcal{L}^\omega$ into a continuum of invariant subclasses, is the entropy per unit of time. The prerequisites from the theory of information are presented in § 1 (the notions of the conditional entropy and of the conditional information are, probably, of wider interest, despite the fact that the exposition closely adheres to the definition of the quantity of information from [7] and the numerous papers developing the ideas of [7]). In § 2 the definition of the characteristics h is presented and a proof of its invariance is given. In § 3, § 4 examples of automorphisms and flows with arbitrary values of h subject to the condition $0 < h \leqslant \infty$ are presented. In the case of automorphisms we deal with examples constructed long ago; in the case of flows constructing examples with finite h is a more delicate problem related to some interesting questions of the theory of Markov processes.

*Reports of the Academy of Sciences of the USSR (DAN SSSR), 1958, V. 119, No. 5, p. 861–864.
†English translation by Nikolai V. Ivanov, 2014.

§ 1
Properties of conditional entropy and of conditional quantity of information

Following [1] we denote by γ the boolean algebra of measurable sets of the space M, considered mod 0. Let $\mathfrak{C}$ be a subalgebra of the algebra γ closed in the metric $\rho(A, B) = \mu((A - B) \cup (B - A))$. This subalgebra generates a well defined mod 0 partition $\xi_{\mathfrak{C}}$ of the space M, defined by the condition that $A \in \mathfrak{C}$ if an only if mod 0 the whole A can be composed of entire members of partition $\xi_{\mathfrak{C}}$. "A canonical system of measures μ_C" is defined on elements C of the partition $\xi_{\mathfrak{C}}$ [1]. For every $x \in C$ we define

$$\mu_x(A \mid \mathfrak{C}) = \mu_C(A \mid C). \tag{2}$$

From the point of view of the probability theory (where every measurable function of elements $x \in M$ is called "a random variable"), the random variable $\mu_x(A \mid \mathfrak{C})$ is nothing else but "the conditional probability" of the event A when the outcome of the "trial" $\mathfrak{C}$ is known [6, ch. 1, § 7].

For three subalgebras $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ of the algebra γ and $C \in \xi_{\mathfrak{C}}$, let

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

where the supremum is taken over finite partitions $M = A_1 \cup A_2 \cup \ldots \cup A_n$, $M = B_1 \cup B_2 \cup \ldots \cup B_n$, such that $A_i \cap A_j = N$, $B_i \cap B_j = N$, $i \neq j$, $A_i \in \mathfrak{A}$, $B_j \in \mathfrak{B}$ (N is the empty set). If $\mathfrak{C}$ is the trivial algebra $\mathfrak{N} = \{N, M\}$, then (3) turns into the definition of unconditional information $I(\mathfrak{A}, \mathfrak{B})$ from Appendix 7 in [7] [1]. The quantity (3) itself is interpreted as "the quantity of information about the results of the trial $\mathfrak{A}$ with respect to the trial $\mathfrak{B}$ when the outcome C of the trial $\mathfrak{C}$ is known". If we do not fix $C \in \xi_{\mathfrak{C}}$, then it is natural to consider the random variable $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$, which is equal to $I_x(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) = I_C(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$ for $x \in C$. In what follows we will deal with its mathematical expectation

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

The definitions of the conditional entropy and of the average conditional entropy $H(\mathfrak{A} \mid \mathfrak{C}) = I(\mathfrak{A}, \mathfrak{A} \mid \mathfrak{C})$, $\mathbf{M}H(\mathfrak{A} \mid \mathfrak{C}) = \int_M H_x(\mathfrak{A} \mid \mathfrak{C})\, \mu(dx)$ do not require any special explanations.

Let us list the properties of the conditional quantity of information and of the conditional entropy which will be needed later. In the case of the unconditional quantity of information and the unconditional entropy the properties (α) and (δ) are well known,

[1] The authors of the note [8] did not paid timely attention to Appendix 7 in [7], which was not included into the Russian translation [9]. The note [8] should start with a reference to this appendix in [7].

the property (ε) for the unconditional quantity of information is the content of Theorem 2 of the note [8]. The proofs of properties (β) and (γ) are easy. Concerning the property (β) one should note only that the similar proposition for the quantity of information instead of the entropy (namely, that $\mathfrak{C} \supseteq \mathfrak{C}'$ implies $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) \geqslant I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}'))$ is wrong. By this reason the lower limit and the symbol $\geqslant$ are present in the property (ζ): the corresponding limit may not exist, and the lower limit in some cases may be bigger than $\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C})$.

(α) $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) \leqslant H(\mathfrak{A} \mid \mathfrak{C})$, the equality is assured if $\mathfrak{B} \supseteq \mathfrak{A}$.

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

(γ) If $\mathfrak{B} \supseteq \mathfrak{B}'$, then $\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) = \mathbf{M}I(\mathfrak{A}, \mathfrak{B}' \mid \mathfrak{C}) + \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C} \vee \mathfrak{B}')$, where $\mathfrak{C} \vee \mathfrak{B}')$ is the minimal closed σ-algebra containing $\mathfrak{C}$ and $\mathfrak{B}'$.

(δ) If $\mathfrak{B} \supseteq \mathfrak{B}'$, then $\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}) \geqslant \mathbf{M}I(\mathfrak{A}, \mathfrak{B}' \mid \mathfrak{C})$.

(ε) If $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \ldots \subseteq \mathfrak{A}_n \ \ldots \ \bigcup_n \mathfrak{A}_n = \mathfrak{A}$, then

$$\lim_{n\to\infty} \mathbf{M}I(\mathfrak{A}_n, \mathfrak{B} \mid \mathfrak{C}) = \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}).$$

(ζ) If $\mathfrak{C}_1 \supseteq \mathfrak{C}_2 \supseteq \ldots \supseteq \mathfrak{C}_n \ \ldots \ \bigcap_n \mathfrak{C}_n = \mathfrak{C}$, then

$$\liminf_{n\to\infty} \ \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}_n) \geqslant \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{C}).$$

§ 2
The definition of the invariant h

We will say that a flow $\{S_t\}$ is quasi-regular (is of the type $\mathcal{R}$) if[2] there is a closed subalgebra γ_0 of the algebra γ, such that its translations $\gamma_t = S_t\gamma_0$ have the following properties:

$$\text{(I)}\ \ \gamma_t \subseteq \gamma_{t'},\ \text{if}\ t \leqslant t'. \quad \text{(II)}\ \ \bigcup_t \gamma_t = \gamma. \quad \text{(III)}\ \ \bigcap_t \gamma_t = \mathfrak{N}.$$

If the flow is interpreted as a stationary random process, then γ_t may be considered as the algebra of events "depending only on the behavior of the process up to the moment of time t". One can easily prove that flows of type $\mathcal{R}$ are transitive, and one can deduce from the results of Plessner [11, 12] that they have homogeneous Lebesgue spectrum. If the multiplicity of the spectrum is equal to ν ($\nu = 1, 2, \ldots, \omega$), then we will say that the flow is a flow of type $\mathcal{R}^\nu$. It is obvious that $\mathcal{R}^\nu \subseteq \mathcal{L}^\nu$, where $\mathcal{L}^\nu$ is the class of flows with

[2]This condition is much weaker that the condition of "regularity" usually used in the theory of random processes. See the end of § 4 for this.

homogeneous Lebesgue spectrum of multiplicity ν. But, perhaps, all $\mathcal{L}^{\nu}$ (and hence all $\mathcal{R}^{\nu}$) except $\mathcal{L}^{\omega}$ ($\mathcal{R}^{\omega}$) are empty, and that $\mathcal{L}^{\omega} = \mathcal{R}^{\omega}$.

T h e o r e m 1. *For the flow* $\{S_t\}$, *if there exist* γ_0 *with the properties* (I), (II), (III), *and if* $\Delta > 0$, *then* $\mathbf{M}H(\gamma_{i+\Delta} \mid \gamma_i) = \Delta h$, *where* h *is a constant such that* $0 < h \leqslant \infty$.

T h e o r e m 2. *For a given flow* $\{S_t\}$ *the constant* h *does not depend on the choice of* γ_0 *with the properties* (I), (II), (III).

Let us outline the proof of Theorem 2. Suppose that the constants $h < \infty$ and h' correspond to two choices γ_0 and γ_0'. By theorem 1 and lemmas (α) and (ε) for every $\varepsilon > 0$ one can find k such that

$$h = \mathbf{M}I(\gamma_{t+1}, \mid \gamma_t) = \mathbf{M}I(\gamma_{t+1}, \gamma \mid \gamma_t) \leqslant \mathbf{M}I(\gamma_{t+1}, \gamma'_{t+k} \mid \gamma_t). \tag{5}$$

By lemma (ζ), from (5) follows that there exists m such that

$$h \leqslant \mathbf{M}I(\gamma_{t+1}, \gamma'_{t+k} \mid \gamma_t \vee \gamma'_s) \text{ for } t - s \geqslant m. \tag{6}$$

It follows from (6) and lemmas (δ), (γ), (α), (β) (which should be used in this order!) that

$$\begin{aligned}
nh &\leqslant \sum_{t=0}^{n-1} \mathbf{M}I(\gamma_{t+1}, \gamma'_{t+k} \mid \gamma_t \vee \gamma'_{-m}) + 2n\varepsilon \leqslant \\
&\leqslant \sum_{t=0}^{n-1} \mathbf{M}I(\gamma_{t+1}, \gamma'_{n+k} \mid \gamma_t \vee \gamma'_{-m}) + 2n\varepsilon = \\
&= \mathbf{M}I(\gamma_n, \gamma'_{n+k} \mid \gamma_0 \vee \gamma'_{-m}) + 2n\varepsilon \leqslant \mathbf{M}H(\gamma'_{n+k} \mid \gamma_0 \vee \gamma'_{-m}) + 2n\varepsilon \leqslant \\
&\leqslant \mathbf{M}H(\gamma'_{n+k} \mid \gamma'_{-m}) + 2n\varepsilon \leqslant (n+k+m)h' + 2n\varepsilon,
\end{aligned}$$

$$h \leqslant \frac{n+k+m}{n} h' + 2\varepsilon. \tag{7}$$

Since $\varepsilon > 0$ and n are arbitrary (and n is chosen after k and m are fixed), (7) implies inequality $h \leqslant h'$. This inequality can be proved in a completely similar manner for the case $h = \infty$ also. The proof of the opposite inequality $h' \leqslant h$ is similar, and this completes the proof of theorem 2.

§ 3
Invariants of automorphisms

If we assume that t runs only over integer values in § 2, then $\{S_t\}$ is uniquely defined by the automorphism $T = S_1$. By Theorems 1 and 2 there is an invariant $0 < h(T) \leqslant \infty$.

It is easy to prove that every automorphism of the type $\mathcal{R}_0$ (the subscript is placed in order to distinguish this case from the case of the flows with continuous time) has the Lebesgue spectrum of countably infinite multiplicity, i.e. among the classes $\mathcal{R}_0^\gamma$ only the class $\mathcal{R}_0^\omega \subseteq \mathcal{L}_0^\omega$ is non-empty. This class splits into subclasses $\mathcal{R}_0^\omega(h)$ according to the values $h(T)$.

T h e o r e m 3. *For every* h, $0 < h \leqslant \infty$, *there exists an automorphism belonging to* $\mathcal{R}_0^\omega(h)$.

The corresponding example are well known and can be constructed, for example, by using the sequence of independent random trials $\mathcal{L}_{-1}, \mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_t, \ldots$ with the probability distribution ξ_t of the trial $\mathcal{L}_t$ given by

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

The space M is assembled from sequences $x = (\ldots, x_{-1}, x_0, x_1, \ldots, x_t, \ldots)$, $x_t = a_1, a_2, \ldots$, and the translation $Tx = x'$ 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

T h e o r e m 4. *For every* h, $0 < h < \infty$, *there exist a flow of the class* $\mathcal{R}_0^\omega(h)$, *i.e. a flow having Lebesgue spectrum of countably infinite multiplicity and with the prescribed value of the constant* h.

The analogy with § 3 naturally suggests the idea of replacing for the proof of Theorem 4 the discrete independent trials by the "processes with independent increments" or by generalized processes "with independent values" [12, 13]. But such an approach leads only to flows of the class $\mathcal{R}_0^\omega(\infty)$ [5]. In order to get finite values h one needs to use some more artificial construction. It is possible to present in this note only an outline of one such construction.

Let us define independent random variables ξ_n, corresponding to all integers n, with the following probability distributions of their values: $\mathbf{P}(\xi_0 = k) = 3 \cdot 4^{-k}$, $k = 1, 2, \ldots$, and $\mathbf{P}(\xi_n = k) = 2^{-k}$, $k = 1, 2, \ldots$, for $n \neq 0$. In the case $\xi_0 = k$, let us consider a random point τ_0 of the t-axis with uniform probability distribution in the interval $-u2^{-k} \leqslant \tau_0 \leqslant 0$, and let us define random points τ_n for $n \neq 0$ by the relation $\tau_{n+1} = \tau_n + u2^{-\xi_n}$.

Let $\varphi(t) = \xi_n$ for $\tau_n \leqslant t < \tau_{n+1}$. It is easy to check that the distribution of the random function $\varphi(t)$ is invariant with respect to the translations $S_t\varphi(t_0) = \varphi(t_0 - t)$. It is easy to calculate that $h\{S_t\} = 6/u$ (during a unit of time one encounters $3/u$ points τ_n in average, and every ξ_n contributes the entropy $\sum\limits_{i=1}^{\infty} k2^{-k} = 2$).

One can get a more graphic description of our random process if we include into the description of its state $\omega(t)$ at moment of time t, in addition to the value $\varphi(t)$, the value $\delta(t) = t - \tau^*(t)$ of the difference between t and the closest to t from the left point

τ_n. If described in this way, our process turns out to be a stationary Markov process. It deserves to be called only "quasi-regular", because, while the corresponding dynamical system is transitive, the value of the difference $f(\omega(t), t) = \tau^*(t) = t - \delta(t)$ is determined up to a dyadic-rational summand by the behaviors of the process realization in any arbitrary far past.

January 21, 1958

References

[1] Rokhlin V. A., Selected topics from the metric theory of dynamical systems (Russian), *Uspehi Matem. Nauk (N.S.)*, Vol. 4 (1949), No. 2(30), 57–128.

[2] Gelfand I. M., Fomin S. V., Geodesic flows on manifolds of constant negative curvature (Russian), *Uspehi Matem. Nauk (N.S.)*, Vol. 7 (1952), No. 1(47), 118–137.

[3] Fomin S. V., On dynamical systems in a space of functions (Russian), *Ukrain. Mat. Žurnal*, Vol. 2, (1950), No. 2, 25–47.

[4] Itô K. Complex multiple Wiener integral, *Jap. J. Math.*, Vol. 22 (1952), 63–86.

[5] Itô K. Spectral type of the shift transformation of differential processes with stationary increments, *Trans. Amer. Math. Soc.*, Vol. 81 (1956), No. 2, 253–263.

[6] Doob J. L. Stochastic processes, *John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London*, 1953, viii+654 pp. Russian translation: *Izdatel'stvo inostrannoi literatury*, 1956.

[7] Shannon C. E., Weaver W. The Mathematical Theory of Communication, *The University of Illinois Press, Urbana, Ill.*, 1949, vi+117 pp.

[8] Gelfand I. M., Kolmogorov A. N., Yaglom A. M., On the general definition of the quantity of information (Russian), *C. R. (Doklady) Acad. Sci. USSR (N.S.)*, Vol. 111 (1956), 745–748.

[9] Shannon C. E., A Mathematical Theory of Communication, *The Bell Systems Technical Journal*, Vol. 27 (1948), 379–423, 623–656, *Russian translation.* In: Teoriya peredachi electricheskih signalov pri nalichii pomeh, 7–87, *Izdatel'stvo inostrannoi literatury*, 1953.

[10] Plessner A. I., Functions of maximal operators (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 23 (1939), No. 4, 327–330.

[11] Plessner A. I., Semiunitary operators (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 25 (1939), 708–710.

[12] Itô K. Stationary random distributions, *Mem. Coll. Sci. Univ. Kyoto. Ser. A. Math.*, Vol. 28 (1954), 209–223.

[13] Gelfand I. M., Generalized random processes (Russian), *Dokl. Akad. Nauk SSSR (N.S.)*, Vol. 100 (1955), 853–856.

Translator: http://nikolaivivanov.com

Addendum to the paper about dynamical systems*†

A. N. Kolmogorov

V. A. Rokhlin pointed out that Theorem 2 of my paper [K] [1] is not correct. Another definition of entropy was suggested in my paper [1] [2]. The shortest way of introducing the notion of entropy was suggested by Y. G. Sinai (see [2]). I would like to attract attention also to the paper [3] of A. G. Koushnirenko, who proved that the entropy of every smooth diffeomorphism of or vector field on a compact smooth manifold with respect to an absolutely continuous invariant measure is finite.

References

[K] Kolmogorov A. N. A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces (Russian), *Dokl. Akad. Nauk SSSR (N.S.)*, 1958, V. 119, No. 5, p. 861–864.

[1] Kolmogorov A. N., On the entropy per unit of time as a metric invariant of automorphisms (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, V. 124 (1959), No. 4, 754–755.

[2] Sinai Y. G., On the concept of entropy for a dynamic system (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 124 (1959), 768–771.

[3] Koushnirenko A. G., An upper bound for the entropy of a classical dynamical system (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 161 (1965), 37–38.

Translator: http://nikolaivivanov.com

*Kolmogorov A. N. Theory of information and theory of algorithms, Nauka Publishing House, Moscow 1987 (304 pp.), p. 275.

†English translation by Nikolai V. Ivanov, 2014.

[1] This paper was reprinted as the paper No. 5 in a volume of selected papers by A. N. Kolmogorov (see the first footnote above); the present note is one of the *Comments and appendices* from this volume. – Translator's note.

[2] In [1], this definition was proved to be correct without using Theorem 2 from [K]. – Translator's note.

On the entropy per unit of time as a metric invariant of automorphisms[*][†]

A. N. Kolmogorov

V. A. Rokhlin pointed out to me that the proof of Theorem 2 from my note [1] implicitly uses the following assumption.

$\mathfrak{A}_1 \supseteq \mathfrak{A}_2 \supseteq \ldots \supseteq \mathfrak{A}_n \supseteq \ldots$, $\bigcap\limits_n \mathfrak{A}_n = \mathfrak{A}$ implies that $\bigcap\limits_n (\mathfrak{B} \vee \mathfrak{A}_n) = \mathfrak{B}$.

V.A. Rokhlin constructed an example showing that Theorem 2 from the note [1] is wrong.[1] Theorems 3 and 4 lose their meaning with the loss of Theorem 2.

In this note I show that in discrete case (i.e. for automorphisms) the number h in examples from § 3 is still an invariant of the corresponding automorphisms. The general definition of this invariant closely follows the definition of entropy of Shannon. Such an approach was developed by me before writing the note [1]. This approach was replaced by the more complicated one, presented in [1], only for the sake of dealing with the more complicated continuous case (i.e. with flows).

---

[*]Reports of the Academy of Sciences of the USSR (DAN USSR), 1959, V. 124, No. 4, p. 754–755.

[†]English translation by Nikolai V. Ivanov, 2014.

[1]With a kind permission of V. A. Rokhlin I will present here his example. Let $\mathfrak{G}^m$ be the additive group of numbers of the form $\alpha m^{-\beta}$ (m is a natural number, α is an integer, β is a non-negative integer. Let us denote by U the automorphism of the group $\mathfrak{G}^6$, acting by division by 6, denote by M the character group of $\mathfrak{G}^6$, and by T the automorphism of M adjoint to U. Subgroups $\mathfrak{G}^2$, $\mathfrak{G}^3$ of the group $\mathfrak{G}^6$ have the following obvious properties

$$\mathfrak{G}^2 \subset \mathfrak{G}^2, \quad \bigvee_n U^n \mathfrak{G}^2 = \mathfrak{G}^6, \quad \bigcap_n U^n \mathfrak{G}^2 = 0,$$

$$\mathfrak{G}^3 \subset \mathfrak{G}^3, \quad \bigvee_n U^n \mathfrak{G}^3 = \mathfrak{G}^6, \quad \bigcap_n U^n \mathfrak{G}^3 = 0.$$

It follows that these subgroups determine subalgebras $\mathfrak{S}^2$, $\mathfrak{S}^3$ of the algebra $\mathfrak{S}$ of measurable sets of the space M, which satisfy the conditions of quasi-regularity. At the same time, $\mathfrak{G}^2$ is a subgroup of $U\mathfrak{G}^2$ of index 3, and $\mathfrak{G}^3$ has index 2 in $U\mathfrak{G}^3$, and hence

$$\mathbf{M}H(T\mathfrak{G}^2 \mid \mathfrak{G}^2) = \lg 3, \quad \mathbf{M}H(T\mathfrak{G}^3 \mid \mathfrak{G}^3) = \lg 2.$$

And indeed, here we have

$$\bigcap_n (\mathfrak{G}^2 \vee T^n \mathfrak{G}^3) \neq \mathfrak{G}^2, \quad \bigcap_n (\mathfrak{G}^3 \vee T^n \mathfrak{G}^2) \neq \mathfrak{G}^3.$$

In this note we keep the definitions and notations of the note [1]. For every closed subalgebra $\mathfrak{A}_0$ of the algebra $\mathfrak{S}$ let

$$\mathfrak{A}_t = S_t\mathfrak{A}_0, \quad \mathfrak{A}_{s,t} = \bigvee_{s<u\leqslant t} \mathfrak{A}_u, \quad \mathfrak{S}(\mathfrak{A}_0) = \bigvee_t \mathfrak{A}_t\,.$$

If the condition

$$\mathfrak{S}(\mathfrak{A}_0) = \mathfrak{S} \tag{A}$$

holds (in the following $\mathfrak{A}_0 \in A$), then a choice of $\mathfrak{A}_0$ leads to a realization of the system of automorphisms S_t as a stationary random process, for which elements of the algebra $\mathfrak{A}_t$ are treated as "the random events observable at the moment of time t". The following theorem is well known (see, for example, [3]).

T h e o r e m 1*. *For every closed subalgebra there exits the limit* $(0 \leqslant h \leqslant \infty)$

$$h(\mathfrak{A}_0) = \lim_{t-s\to\infty} H(\mathfrak{A}_{s,t}).$$

L e m m a 1. *If* $h(\mathfrak{A}_0) < \infty$ *and* $\mathfrak{A}'_0 \in A$, *then*

$$h(\mathfrak{A}_0) \leqslant h(\mathfrak{A}'_0).$$

For the proof we note that by (A)

$$\bigvee_n \mathfrak{A}'_{-n,n} = \mathfrak{S}$$

and that in view of (ε) from [1] for every $\varepsilon > 0$ there is n such that

$$I(\mathfrak{A}_0, \mathfrak{A}_{-n,n}) \geqslant H(\mathfrak{A}_0) - \varepsilon,$$

i.e.

$$\mathbf{M}H(\mathfrak{A}_0 \mid \mathfrak{A}'_{-n,n}) = H(\mathfrak{A}_0) - I(\mathfrak{A}_0, \mathfrak{A}'_{-n,n}) \leqslant \varepsilon,$$

$$\mathbf{M}H(\mathfrak{A}_{s,t} \mid \mathfrak{A}'_{s-t,t+n}) \leqslant \sum_{s<u\leqslant t} \mathbf{M}H(\mathfrak{A}_u \mid \mathfrak{A}'_{u-n,u+n}) \leqslant (t-s)\varepsilon.$$

Therefore

$$H(\mathfrak{A}_{s,t}) \leqslant I(\mathfrak{A}_{s,t}, \mathfrak{A}'_{s-n,t+n}) + \mathbf{M}H(\mathfrak{A}_{s,t} \mid \mathfrak{A}'_{s-n,t+n}) \leqslant$$
$$\leqslant H(\mathfrak{A}'_{s-n,t+n}) + (t-s)\varepsilon,$$

or, after dividing by $t-s$ and taking the limit for $t-s \to \infty$:

$$h(\mathfrak{A}_0) \leqslant h(\mathfrak{A}'_0) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves the lemma.

The following theorem immediately follows from Lemma 1:

T h e o r e m 2*. *If* $\mathfrak{A}_0 \in A$, $\mathfrak{A}'_0 \in A$ *and* $h(\mathfrak{A}_0)$, $h(\mathfrak{A}'_0)$ *are finite, then*

$$h(\mathfrak{A}_0) = h(\mathfrak{A}'_0).$$

Theorems 1* and 2* are similar to Theorems 1 and 2 of the note [1]. If we set

$$h_1(T) = \inf_{\mathfrak{A}_0 \in A} h(\mathfrak{A}_0),$$

then we see that h_1 is equal to ∞ if $h(\mathfrak{A}_0) = \infty$ for all $\mathfrak{A}_0 \in A$; but if there is an $\mathfrak{A}_0 \in A$ with finite $h(\mathfrak{A}_0)$, then the value $h(\mathfrak{A}_0)$ for all such $\mathfrak{A}_0$ is the same and $h_1(T)$ is equal to the common value of all finite $h(\mathfrak{A}_0)$. Theorem 2* (and Lemma 1 directly in the case of infinite h) easily implies that in the examples of § 3 of the note [1] with

$$h = -\sum p_i \log p_i$$

the invariant h_1 is equal to h.

Remarks added in proofs.

1. Recently Y. Sinai managed to transfer the definition of the entropy introduced by him in [2] to the flows in such a way that the entropy is invariant with respect to "isomorphisms modulo zero" and flows constructed in § 4 of my note [1] have the entropy $6/u$, as was claimed in [1], and therefore are not isomorphic for different values u.

2. Concerning the very idea of using the notions of the theory of information for construction of invariants of automorphisms and flows, one should mention that this idea without any specific results is also contained in the diploma paper (an analogue of the master thesis – translator's note) of D. Z. Arov, defended at the Odessa State University in the spring of 1957.

Received
November 25, 1958

References

[1] Kolmogorov A. N. A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, 1958, V. 119, No. 5, p. 861–864.

[2] Sinai Y. G. On the concept of entropy for a dynamic system (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 124 (1959), 768–771.

[3] Khinchin A. Y. On the basic theorems of information theory (Russian), *Uspehi Mat. Nauk (N.S.)*, Vol. 11 (1956), No. 1(67), 17–75.

Translator: http://nikolaivivanov.com

On the notion of entropy of a dynamical system*(an excerpt)†‡§

Y. G. Sinai

(Submitted by academician A. N. Kolmogorov at November 25, 1958)

§ 1. In this note we present a definition of entropy suitable for arbitrary automorphisms of Lebesgue spaces. Theorem 1, proved right after the definition, suggests how to compute the entropy in some cases. This theorem serves as a basis for producing new metric invariants of some ergodic automorphisms of compact commutative groups.

§ 2. Let M be a Lebesgue space, S be its σ-algebra if measurable sets and μ be the measure of M, $\mu(M) = 1$; and let T be an arbitrary automorphism of this space [2]. A finite partition $A = \{A_1 \dots A_n\}$ of the space M is understood as a presentation of M as a sum of disjoint sets[1] $M = \bigcup\limits_{i=1}^{n} A_i$. The partition $T^k A$ is understood as the partition formed by set $\{T^k A_i\}$, $i = 1, \dots, n$.

The entropy of an arbitrary finite partition A is defined by the well known formula

$$h(A) = -\sum_{i=1}^{n} \mu(A_i) \log \mu(A_i).$$

This formula allows us to compute the value of the entropy of the partition $A \vee TA \vee \dots \vee T^n A$ defined by all possible sets of the form $A_{i_0} \cap TA_{i_1} \cap \dots \cap T^n A_{i_n}$. It follows from the general theorems of the theory of information [4] that for every finite partition A there exist the limit

$$\lim_{k\to\infty} \frac{h(A \vee TA \vee \dots \vee T^k A)}{k} = h_T(A).$$

*C. R. (Doklady) Acad. Sci. URSS (N. S.), Vol. 124 (1959), No. 4, 768–771.

†This excerpt includes only § 1, which is a four lines long introduction to the paper, and § 2, which is devoted to the author's version of the definition of entropy (this version quickly became the standard one and was considered as the best one by A. N. Kolmogorov). The § 3, which is longer and is devoted to a computation of the entropy of ergodic automorphisms of a 2-dimensional torus, is not included.

‡For a free translation of the whole paper by the author himself, see Y. G. Sinai, *Selecta, Volume I: Ergodic Theory and Dynamical Systems*, Springer Science+Business Media, LLC, 2010, 3–8.

§English translation by Nikolai V. Ivanov, 2014. This is a faithful translation of an excerpt from the original Russian publication and differs from author's free translation cited in the previous footnote. Some misprints in formulas in the original Russian paper are corrected here, but were left uncorrected in author's translation.

[1] All sets considered are assumed to be measurable.

D e f i n i t i o n. The upper bound of $h_T(A)$ over all finite partitions A will be called the e n t r o p y o f t h e a u t o m o r p h i s m T: $h_T = \sup_A h_T(A)$.

Let us consider two arbitrary partitions $A = \{A_1 \dots A_k\}$ and $B = \{B_1 \dots B_l\}$. Let us assume that the sets B_i belong to the closed σ-algebra generated by sets $\{T^n A_j,\ 1 \leqslant j \leqslant k,\ -\infty < n < \infty\}$. Then the following theorem holds.

T h e o r e m 1. *For every two partitions* A *and* B *satisfying this assumption, the inequality* $h_T(B) \leqslant h_T(B)$ *holds.*

Let us list some well known properties of the entropy, which are basic for our proof. Let K, L, M be arbitrary finite partitions. Then

$$\begin{aligned}
&1) && h(K \vee L) = h(K) + h(L \mid K);\\
&2\alpha) && h(K) \leqslant h(K \vee L) \leqslant h(K) + h(L);\\
&2\beta) && h(K \vee L \mid m) \leqslant h(K \mid M) + h(L \mid M);\\
&3) && h(K \mid L) \geqslant h(K \mid M),
\end{aligned}$$

if the elements of the partition L can be presented as sums of elements of the partition M.

P r o o f o f T h e o r e m. By properties 1) and 2α)

$$\begin{gathered}
h(B \vee TB \vee \dots \vee T^r B) \leqslant \\
\leqslant h(B \vee TB \vee \dots \vee T^r B \vee T^{-n} A \vee \dots T^n A \vee \dots \vee T^{n+r} A) = \\
= h(T^{-n} A \vee \dots \vee T^{n-r} A) + h(B \vee TB \vee \dots \vee T^r B \mid T^{-n} A \vee \dots \vee T^{n+r} A).
\end{gathered} \tag{1}$$

Next, by properties 2β and 3)

$$\begin{gathered}
h(B \vee TB \vee \dots \vee T^r B \mid T^{-n} A \vee \dots \vee T^{n+r} A) \leqslant \\
\leqslant \sum_{i=0}^{r} h(T^i B \mid T^{-n} A \vee \dots \vee T^{n+r} A) \leqslant \\
\leqslant \sum_{i=0}^{r} h(T^i B \mid T^{-n+i} A \vee \dots \vee T^{n+i} A) = (r+1)\, h(B \mid T^{-n} A \vee \dots \vee T^n A).
\end{gathered} \tag{2}$$

By using our assumption about partitions A and B, it is easy to show that for every $\varepsilon > 0$ the inequality $h(B \mid T^{-n} A \vee \dots \vee T^n A) < \varepsilon$ holds for sufficiently large n. By dividing both parts of the inequality (1) and using our last statement and the inequality (2), we get

$$\frac{h(B \vee TB \vee \dots \vee T^r B)}{r} \leqslant \frac{h(TA \vee \dots \vee T^{2n+r} A)}{2n+r} \frac{2n+r}{r} + \varepsilon \frac{r+1}{r}.$$

Since n depends only on ε, by taking the limit for $r \to \infty$ and noticing that ε is arbitrary, we get the required result.

C o r o l l a r y. If the partition A is such that the closed σ-algebra generated by the sets $\{T^k A_i\}$, $-\infty < k < \infty$, $1 \leqslant i \leqslant k$, is equal to S, then $h_T = h_T(A)$.

T h e o r e m 2. *If the partition* A *is such that the minimal* σ*-algebra contag sets* $\{T^k A_i\}$ *for* $k \geqslant 0$ *is equal to* S*, then* $h_T = 0$.

P r o o f is based on the fact that $h(A \mid TA \vee \ldots \vee T^n A) \to 0$ for $n \to \infty$ under our assumptions.

T h e o r e m 3. *For every automorphism* T *the equality* $h_{T^k} = |k| h_T$ *holds.*

P r o o f. For every partition A we have

$$\frac{h(A \vee \ldots \vee T^{nk} A)}{n|k|} \geqslant \frac{h(A \vee T^k \vee \ldots \vee T^{nk} A)}{n|k|},$$

or, after the limit for $n \to \infty$, $h_T(A) \geqslant \frac{1}{|k|} h_{T^k}(A)$.

The last inequality implies that $h_T \geqslant \frac{1}{|k|} h_{T^k}$. Now, let h' be any number less than h_T. Then there is a partition A such that $h_T(A) \geqslant h'$. Consider the partition $B = A \vee TA \vee \ldots \vee T^{k-1} A$. Obviously, $h(B \vee T^k B \vee \ldots \vee T^{k(n-1)} B) = h(A \vee TA \vee \ldots \vee T^{kn-1} A)$, and

$$h_{T^k} \geqslant \frac{h(B \vee T^k B \vee \ldots \vee T^{k(n-1)} B)}{n} = \frac{h(A \vee TA \vee \ldots \vee T^{kn-1} A)}{|k|n} |k|.$$

Taking the limit for $n \to \infty$ leads to $h_{T^k} \geqslant h'|k|$. Since $h' \leqslant h_T$ is arbitrary, this proves the theorem.

References

[1] Kolmogorov A. N. A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces (Russian), *Dokl. Akad. Nauk SSSR (N.S.)*, 1958, V. 119, No. 5, p. 861–864.

[2] Rokhlin V. A., Selected topics from the metric theory of dynamical systems (Russian), *Uspehi Matem. Nauk*, Vol. 4 (1949), No. 2(30), 57–128.

[3] Rokhlin V. A., On endomorphisms of compact commutative groups, (Russian), *Izvest. Acad. Nauk (N.S.)*, Vol. 13 (1949), 329–340.

[4] Khinchin A. Y., On the basic theorems of the theory of information, (Russian), *Uspehi Matem. Nauk*, Vol. 11 (1956), No. 1, 17–75.

Translator: http://nikolaivivanov.com

Flows with finite entropy[*†]

Y. G. Sinai

(Submitted by academician A. N. Kolmogorov at January 16, 1959)

§ 1. Let us consider a flow $(M, \mathfrak{S}, S^t)$, where M is a Lebesgue space with the measure μ; $\mathfrak{S}$ is the σ-algebra of its measurable sets, and S^t, $-\infty < t < \infty$, is a group of measure preserving transformations acting on this space (see [2]). According to [3], for individual automorphisms from this group one can introduce the notion of entropy $h_{S^\Delta} = h_S(\Delta)$. In all known examples the dependence of this function on Δ turns out to be linear, but there is no proof of this fact in the general case. By this reason the entropy of the flow S^t is defined as the supremum

$$\sup_{\Delta > 0} \frac{h_{S^\Delta}}{\Delta} = h_S.$$

For the computations of entropy of flows, it is convenient to use the following theorem, which holds for every automorphism T.

T h e o r e m. *Let* $\{g_t\}$ *be a sequence of partitions such that* $g_k \leqslant g_{k+1}$, $\prod_k \prod_{n=-\infty}^{\infty} \{T^n g_k\} = \varepsilon$, *where* ε *is the partition into individual points, and* $h(g_k) < \infty$. *Then*

$$h_T = \lim_{k \to \infty} h_T(g_k).$$

The proof of this theorem is similar to the proof of Theorem 1 in [3].

It is proved in this note that there exist transitive flows with Lebesgue spectrum of countably infinite multiplicity and arbitrary finite $h_S > 0$. The examples of flows with $h_S = \infty$ are trivial. The example from § 3 was published in A. N. Kolmogorov's paper [1]. The examples from § 2 are also due to Kolmogorov.

§ 2. Let us consider a Markov process having as its phase space two segments A_1B_1 and A_2B_2 of length α and β respectively. In each of the segments the deterministic uniform motion from the left to the right with unit speed takes place. A jump to A_1 or A_2 with probabilities $1/2$, $1/2$ happens at the right end (B_1 or B_2) of each of the segments. It is known that a local description of the motion together with a given initial distribution uniquely defines a measure on the space of trajectories of a Markov process.

*C. R. (Doklady) Acad. Sci. URSS (N. S.), Vol. 125 (1959), No. 6, 1200–1202.
†English translation by Nikolai V. Ivanov, 2014.

If we take the uniform distribution on both segments as the initial distribution, then the resulting Markov process will be stationary.

In order to compute the entropy of the flow constructed in this way one needs to compute the entropy of the corresponding automorphisms S^{Δ}. Let us partition A_1B_1 into segments of length $\alpha/2^k$. Let us partition A_2B_2 into segments of the same length $\alpha/2^k$ with only possible the exception of the last one. Altogether we will get $2^k[1+\frac{\beta}{\alpha}+1]$ segments I_i^k. Let g^k be the partition of the whole space into the sets G_i^k of trajectories located at time moment 0 in the interval I_i^k. It is easy to see that for $\Delta < \min(\alpha, \beta)$ the sequence of partitions g^k satisfies the conditions of Theorem.

Let us look what happens with the sets G_i^k when transformations $S^{\Delta n}$ are applied to them. In other words, we need to consider how the sets G_i^k are divided into parts according to the possible positions of the trajectories of the Markov process in the moments of time $0, \Delta, \dots, n\delta, \dots$. In the case of trajectories with deterministic motion during the period of time from 0 till Δ each individual set G_i^k is divided into no more than two parts. In the case of sets G_i^k consisting of trajectories having a jump during the period of time from 0 till Δ, each set G_i^k splits into the sum of two sets $G_{i_1}^k$ and $G_{i_2}^k$ according to which of the segments, A_1B_1 or A_2B_2, the trajectory will belong at the moment of time Δ; both of these two sets have the conditional measure equal to $1/2$. The further analysis of our partition can be done independently for each of the sets $G_{i_1}^k$ and $G_{i_2}^k$. Some of the sets G_i^k may happen to be such that some part of the trajectories has a jump during the period of time from 0 till Δ, and the other part is moving deterministically. The partition in this case is the combination of two previous cases and does not differ from them.

After n steps the set G_i^k will be divided into separate subsets depending on the number of the jumps of a trajectory. The conditional measure of such a subset is equal to $2^{-\nu}$, where ν is the number of jumps. In addition, each subset with a fixed number of jumps will be subdivided into no more than $(n+1)$ parts depending on into which of the segments I_i^k the trajectory will get after the next step.

In order to complete the computation of the entropy it remains to point out that the law of large numbers implies that, given an arbitrarily small ε, then with probability higher that 1ε we have

$$\frac{2}{\alpha+\beta}(1-\delta) < \frac{\nu}{n\Delta} < \frac{2}{\alpha+\beta}(1+\delta)$$

for every δ and all sufficiently bin n. Therefore

$$(1-\varepsilon)\frac{2}{\alpha+\beta}(1-\delta) \leqslant \frac{H(\prod_{i=0}^{n} T^{i\Delta} g^k)}{n\Delta} \leqslant$$

$$\leqslant \varepsilon \cdot 2^k \left(2+\frac{\beta}{\alpha}\right) + \frac{2}{\alpha+\beta}(1+\delta) + \frac{\log(n+1)}{n\Delta} - \frac{\varepsilon \log \varepsilon + (1-\varepsilon)\log(1-\varepsilon)}{n\Delta}.$$

By taking the limit for $n \to \infty$, and then taking the limit for $\varepsilon, \mathtt{delta} \to 0$, we conclude that

$$h_{S^\Delta}(g^k) = \frac{2\Delta}{\alpha + \beta}.$$

It should be noted that the equality

$$h_{S^\Delta}(g^k) = \frac{2\Delta}{\alpha + \beta}.$$

holds for every k. If α and β are non-commensurable, then the paron g^k is a generating partition for every k, and we get an example of a finitely generated automorphism which can be embedded into a flow.

Speaking about the spectrum, it is not difficult to prove that in the case of non-commensurable α and β the spectrum will be a Lebesgue spectrum of countably infinite multiplicity[1]. Therefore, the considered flows provide us with examples of dynamical systems which are spectrally isomorphic, but metrically non-isomorphic.

§ 3. Let us consider the example of a flow published in [1]. This example is a Markov process having as the phase space a countable set of segments $\Gamma_i = \{A_i, B_i\}$ of length $u/2^i$. On every segment we have the deterministic motion from the left to the right up to the right end of the segment; at the right end a jump to the point A_j with probability 2^{-j} occurs. The stationary distribution is the following: the probability of choosing l-th segment is $3/4^l$, and the distribution on the l-th segment is uniform.

Let us divide each segment into the segments I_i^k of length $u/2^k$, if such a subdivision is possible. Let us left other segments unchanged. Let h^k be the partition into trajectories which are in segments I_i^k in the initial moment of time. Let us denote by $M_{i_1 i_2 \dots i_l}$ the set of trajectories which spend the period of time from $-\Delta$ till 0 entirely in the segments Γ_{i_1}, $\Gamma_{i_2}, \dots, \Gamma_{i_l}$. Obviously, this is possible only if $\frac{u}{2^{i_1}} + \frac{u}{2^{i_2}} + \dots + \frac{u}{2^{i_l}} < \Delta$. Let M_0 be the set of trajectories having no more than one jump during the period of time from $-\Delta$ till 0. Let us denote by m the partition of the space into sets $M_{i_1 i_2 \dots i_l}$ and M_0. Let $g^k = m \cdot h^k$. Then the conditions of Theorem 1 hold for the sequence of partitions g^k.

The partition into the sets G_i^k is constructed completely similarly to § 2 by taking into account what happens in time. If we fix the order of jumps, we will get a partition of G_i^k into sets with the conditional measure equal to

$$p = 2^{-\nu_1 - \nu_2 - \dots - \nu_k},$$

where $\nu_1, \dots, \nu_k$ are the numbers of the segments, to which the trajectory belongs after the next jumps. The law of large numbers implies that with probability $1 - \varepsilon$ the inequality

$$\frac{3}{u}(1 - \delta) \leqslant \frac{k}{n\Delta} \leqslant \frac{3}{u}(1 + \delta)$$

[1]This result was proved also by Girsanov.

holds for the number of jumps k, and the inequality

$$\frac{6}{u}(1-\delta_1) \leqslant -\frac{\log p}{n\Delta} \leqslant \frac{6}{u}(1+\delta_1)$$

holds for the probabilities of the corresponding jumps.

Every set with a fixed order of jumps is subdivided into no more than $k+n+1$ part according to the position of the trajectory at the moments $i\Delta$. Therefore,

$$(1-\varepsilon)\frac{6}{u}(1-\delta_1) \leqslant \frac{H\left(\prod_{i=0}^{n} S^{i\Delta} g^k\right)}{n\Delta} \leqslant$$

$$\leqslant \varepsilon f(k) + \frac{6}{u}(1+\delta_1) + \frac{\log n\left(1+\Delta\frac{3}{u}(1+\delta)\right)}{n\Delta} - \frac{\varepsilon\log\varepsilon + \log(1-\varepsilon)}{n\Delta},$$

where $f(k)$ is the entropy of the initial distribution g^k. By taking the limit for $n \to \infty$ and ε, δ, $\delta_1 \to 0$, we get $\frac{1}{\Delta}h_{S^\Delta}(g^k) = \frac{6}{u}$. It follows that $h_S = 6/u$.

Finally, my pleasant duty is to express my gratitude to A. N. Kolmogorov for his guidance during my work which lead to this paper.

References

[1] Kolmogorov A. N. A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces (Russian), *Dokl. Akad. Nauk SSSR (N.S.)*, 1958, V. 119, No. 5, p. 861–864.

[2] Rokhlin V. A., Selected topics from the metric theory of dynamical systems (Russian), *Uspehi Matem. Nauk (N.S.)*, Vol. 4 (1949), No. 2(30), 57–128.

[3] Sinai Y. G., On the concept of entropy for a dynamic system (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 124 (1959), No. 4, 768–771.

Translator: http://nikolaivivanov.com

A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces*†‡

A. N. Kolmogorov

This paper is a revised version of my note [1] with the same title, the Theorem 2 of which is wrong. In this version both the statement and the proof of Theorem 2 are replaced by completely different ones. This revision was prepared for the anniversary issue of the Proceedings of the Steklov Mathematical Institute of the Academy of Sciences.

It is well known that a significant part of the metric theory of dynamical systems can be presented as an abstract theory of "flows" $\{S_t\}$ on "Lebesgue spaces" M with a measure μ in a form, invariant with respect to "isomorphisms modulo zero" (see the expository paper [2] by V. A. Rokhlin, the definitions and notations of which are adopted in the following exposition). We will assume that the measure on M is normalized by the condition

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

and is non-trivial (i.e., we assume that there exist a set $A \subset M$ with $0 < \mu(A) < 1$). Many examples of transitive automorphisms and transitive flows with so-called Lebesgue spectrum of countably infinite multiplicity are known (for automorphisms see [2, § 4], for flows [3–6]). From the spectral point of view we have here only one type of automorphisms $\mathcal{L}_0^\omega$ and only one type of flows $\mathcal{L}^\omega$. The question if all automorphisms of type $\mathcal{L}_0^\omega$ (respectively, all flows of type $\mathcal{L}^\omega$) are metrically isomorphic to each other mod 0 remained open up to now. We show in Sections 3–4 that the answer to this question negative both in the case of automorphisms and in the case of flows. The new invariant, which allows to split the class of automorphisms $\mathcal{L}_0^\omega$ and the class of flows $\mathcal{L}^\omega$ into a continuum of invariant subclasses, is the entropy per unit of time. The prerequisites from the theory of information are presented in Section 1 (the notions of the conditional entropy and of the conditional information are, probably, of wider interest, despite the

*Proceeding of the Mathematical Institute of AN USSR, 1985, V. 169, p. 94–98.

†English translation by Nikolai V. Ivanov, 2014.

‡Another English translation was published in *Proceedings of the Steklov Institute of Mathematics*, Vol. 169 (1986), No. 4, 97–102 (AMS). The present translator hopes that the new translation is more faithfully follows the Russian original and the original (not translated into English before) 1958 version of this paper. The AMS translation was not used while preparing this one. With the exception of § 2, the two versions are almost identical, and so are the translations of both versions by the present translator.

fact that the exposition closely adheres to the definition of the quantity of information from [8] and the numerous papers developing the ideas of [8]). In Section 2 the definition of the characteristics h is presented and a proof of its invariance is given. In Sections 3–4 examples of automorphisms and flows with arbitrary values of h subject to the condition $0 < h \leqslant \infty$ are presented. In the case of automorphisms we deal with examples constructed long ago; in the case of flows constructing examples with finite h is a more delicate problem related to some interesting questions of the theory of Markov processes.

1. Properties of conditional entropy and of conditional quantity of information

Following [2] we denote by γ the boolean algebra of measurable sets of the space M, considered mod 0. Let $\mathfrak{L}$ be a subalgebra of the algebra γ closed in the metric $\rho(A, B) = \mu((A - B) \cup (B - A))$. This subalgebra generates a well defined mod 0 partition $\xi_{\mathfrak{L}}$ of the space M, defined by the condition that $A \in \mathfrak{L}$ if an only if mod 0 the whole A can be composed of entire members of partition $\xi_{\mathfrak{L}}$. "A canonical system of measures μ_C" is defined on elements C of the partition $\xi_{\mathfrak{L}}$ [2]. For every $x \in C$ we define

$$\mu_x(A \mid \mathfrak{C}) = \mu_C(A \mid C). \tag{2}$$

From the point of view of the probability theory (where every measurable function of elements $x \in M$ is called "a random variable"), the random variable $\mu_x(A \mid \mathfrak{C})$ is nothing else but "the conditional probability" of the event A when the outcome of the "trial" $\mathfrak{L}$ is known [7, ch. 1, § 7].

For three subalgebras $\mathfrak{A}$, $\mathfrak{B}$, $\mathfrak{C}$ of the algebra γ and $C \in \xi_{\mathfrak{L}}$, let

$$I_C(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) = \sup \sum_{i,j} \mu_x(A_i \cap B_j) \log \frac{\mu_x(A_i \cap B_j)}{\mu_x(A_i)\mu_x(B_j)}, \tag{3}$$

where the supremum is taken over finite partitions $M = A_1 \cup A_2 \cup \ldots \cup A_n$, $M = B_1 \cup B_2 \cup \ldots \cup B_n$, such that $A_i \cap A_j = N$, $B_i \cap B_j = N$, $i \neq j$, $A_i \in \mathfrak{A}$, $B_j \in \mathfrak{B}$ (N is the empty set). If $\mathfrak{L}$ is the trivial algebra $\mathfrak{N} = \{N, M\}$, then (3) turns into the definition of unconditional information $I(\mathfrak{A}, \mathfrak{B})$ from Appendix 7 in [8] [1]. The quantity (3) itself is interpreted as "the quantity of information about the results of the trial $\mathfrak{A}$ with respect to the trial $\mathfrak{B}$ when the outcome C of the trial $\mathfrak{L}$ is known". If we do not fix $C \in \xi_{\mathfrak{L}}$, then it is natural to consider the random variable $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L})$, which is equal to $I_x(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) = I_C(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L})$ for $x \in C$. In what follows we will deal with its mathematical expectation

$$\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) = \int_M I_x(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) \, \mu(dx). \tag{4}$$

[1] The authors of the note [9] did not paid timely attention to Appendix 7 in [8], which was not included into the Russian translation [10]. The note [9] should start with a reference to this appendix in [8].

The definitions of the conditional entropy and of the average conditional entropy $H(\mathfrak{A} \mid \mathfrak{L}) = I(\mathfrak{A}, \mathfrak{A} \mid \mathfrak{L})$, $\mathbf{M}H(\mathfrak{A} \mid \mathfrak{L}) = \int_M H_x(\mathfrak{A} \mid \mathfrak{L})\, \mu(dx)$ do not cause any special difficulties.

Let us list the properties of the conditional quantity of information and of the conditional entropy which will be needed later. In the case of the unconditional quantity of information and the unconditional entropy the properties (α) and (δ) are well known, the property (ε) for the unconditional quantity of information is the content of Theorem 2 of the note [9]. The proofs of properties (β) and (γ) are easy. Concerning the property (β) one should note only that the similar proposition for the quantity of information instead of the entropy (namely, that $\mathfrak{L} \supseteq \mathfrak{L}'$ implies $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) \geqslant I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}')$) is wrong. By this reason the l o w e r l i m i t and the s y m b o l $\geqslant$ are present in the property (ζ): the corresponding limit may not exist, and the lower limit in some cases may be bigger than $\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L})$.

(α) $I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) \leqslant H(\mathfrak{A} \mid \mathfrak{L})$, the equality is assured if $\mathfrak{B} \supseteq \mathfrak{A}$.

(β) If $\mathfrak{L} \supseteq \mathfrak{L}'$, then $H(\mathfrak{A} \mid \mathfrak{L}) \leqslant H(\mathfrak{A} \mid \mathfrak{L}')$, mod 0.

(γ) If $\mathfrak{B} \supseteq \mathfrak{B}'$, then $\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) = \mathbf{M}I(\mathfrak{A}, \mathfrak{B}' \mid \mathfrak{L}) + \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L} \vee \mathfrak{B}')$, where $\mathfrak{L} \vee \mathfrak{B}')$ is the minimal closed σ-algebra containing $\mathfrak{L}$ and $\mathfrak{B}'$.

(δ) If $\mathfrak{B} \supseteq \mathfrak{B}'$, then $\mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}) \geqslant \mathbf{M}I(\mathfrak{A}, \mathfrak{B}' \mid \mathfrak{L})$.

(ε) If $\mathfrak{A}_1 \subseteq \mathfrak{A}_2 \subseteq \ldots \subseteq \mathfrak{A}_n \ \ldots \ \bigcup\limits_n \mathfrak{A}_n = \mathfrak{A}$, then $\lim_{n\to\infty} \mathbf{M}I(\mathfrak{A}_n, \mathfrak{B} \mid \mathfrak{L}) =$
$= \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L})$.

(ζ) If $\mathfrak{L}_1 \supseteq \mathfrak{L}_2 \supseteq \ldots \supseteq \mathfrak{L}_n \ \ldots \ \bigcap\limits_n \mathfrak{L}_n = \mathfrak{L}$, then $\liminf_{n\to\infty} \ \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L}_n) \geqslant$
$\geqslant \mathbf{M}I(\mathfrak{A}, \mathfrak{B} \mid \mathfrak{L})$.

2. The definition of the invariant h

We will say that a flow $\{S_t\}$ is quasiregular (is of the type $\mathfrak{R}$) if [2] there is a closed subalgebra γ_0 of the algebra γ, such that its translations $\gamma_t = S_t\gamma_0$ have the following properties: (I) $\gamma_t \subseteq \gamma_{t'}$, if $t \leqslant t'$. (II) $\bigcup\limits_t \gamma_t = \gamma$. (III) $\bigcap\limits_t \gamma_t = \mathfrak{N}$.

If the flow is interpreted as a stationary random process, then γ_t may be considered as the algebra of events "depending only on the behavior of the process u p t o t h e m o m e n t o f t i m e t". One can easily prove that flows of type $\mathcal{R}$ are transitive, and one can deduce from the results of Plessner [12, 19] that they have homogeneous Lebesgue spectrum. If the multiplicity of the spectrum is equal to ν ($\nu = 1, 2, \ldots, \omega$), then we will say that the flow is a flow of type $\mathcal{R}^\nu$. It is obvious that $\mathcal{R}^\nu \subseteq L^\nu$, where L^ν is the class of flows with homogeneous Lebesgue spectrum of multiplicity ν. It was shown in [13] that class of flows $\mathcal{R}^\nu$ with finite ν is empty. At the same time $\mathcal{R}^\infty \neq L^\infty$. For example, the

[2] This condition is much weaker that the condition of "regularity" usually used in the theory of random processes. See the end of § 4 for this.

horocyclic flow on a compact closed surface of constant negative curvature has Lebesgue spectrum of countably infinite multiplicity (see [14]) and is not quasiregular (see [15]).

T h e o r e m 1. *For the flow* $\{S_t\}$, *if there exist* γ_0 *with the properties* (I), (II), (III), *and if* $\Delta > 0$, *then* $\mathbf{M}H(\gamma_{i+\Delta} \mid \gamma_i) = \Delta h(\gamma_0)$, *where* $h(\gamma_0)$ *is a constant such that* $0 < h(\gamma_0) \leqslant \infty$.

The proof of Theorem 1 uses only the property (I) of quasiregular subalgebra. For a general flow $\{S_t\}$, we will call subalgebra γ_0 monotone, if $S_t(\gamma_0)) \supseteq \gamma_0$ for $t > 0$. For monotone subalgebras the conclusion of Theorem 1 still holds.

The paper [1] included also Theorem 2, which is wrong, as was pointed out to me by Rokhlin (see [16]). The rest of this section differs from the corresponding part of [1].

It is simpler to deal with automorphisms. It is obvious how to transfer the definitions of a quasiregular and a monotone subalgebra to this case. Some simple examples of monotone subalgebras can be constructed as follows. Let $\mathfrak{A}_0$ be an arbitrary finite subalgebra, $\mathfrak{A}^n_m = \bigcup\limits_{m<k\leqslant n} T^k\mathfrak{A}_0$ for arbitrary $-\infty \leqslant m < n < \infty$. Then $\gamma_0 = \mathfrak{A}^0_{-\infty}$ will be a monotone subalgebra. We will call such monotone subalgebras finitely generated.

T h e e n t r o p y o f a n a u t o m o r p h i s m T is defined as the number $h_0 = \sup H(T\gamma_0 \mid \gamma_0)$, where the supremum is taken of all monotone subalgebras.

Let

1) $h = \sup H(T\gamma_0 \mid \gamma_0)$, where the supremum is taken over all finitely generated monotone subalgebras;

2) $h_1 = \sup H(T\gamma_0 \mid \gamma_0)$, where the supremum is taken over all monotone subalgebras such that $\bigcup\limits_n T^n\gamma_0 = \gamma$;

3) $h_2 = \sup H(T\gamma_0 \mid \gamma_0)$, where the supremum is taken over all monotone subalgebras such that $\bigcap\limits_n T^n\gamma_0 = \mathfrak{N}$.

Each of the numbers h, h_0, h_1, h_2 is, by its definition, a metric invariant of the automorphism. The invariant h was introduced in [17] and now serves as the most widely used definition of the entropy of an automorphism.

T h e o r e m 2. *In the case of transitive automorphisms* $h = h_0 = h_1 = h_2$.

P r o o f. Let γ_0 be a monotone subalgebra. Consider an increasing sequence of finite subalgebras $\mathfrak{A}^{(n)} \subseteq \gamma_0$, $\bigcup\limits_n \mathfrak{A}^{(n)} = \gamma_0$. Since γ_0 is monotone, we have $(\mathfrak{A}^{(n)})^0_{-\infty} \subseteq \gamma_0$, $\bigcup\limits_n (\mathfrak{A}^{(n)})^0_{-\infty} = \gamma_0$. By applying properties (ε) and (β), we get

$$H(T\gamma_0 \mid \gamma_0) = \lim_{n\to\infty} H(T\mathfrak{A}^{(n)} \mid \gamma_0) \leqslant \lim_{n\to\infty} H(T\mathfrak{A}^{(n)} \mid (\mathfrak{A}^{(n)})^0_{-\infty}).$$

Since all monotone subalgebras $(\mathfrak{A}^{(n)})^0_{-\infty}$ are finitely generated, this immediately implies that $h = h_0 = h_2$.

The proof of the equality $h = h_2$ follows from a result of [18] to the effect that for every finitely generated monotone subalgebra γ_0 there exists a finite subalgebra $\mathfrak{A}_0 \subseteq \gamma_0$, such that all translations $T^n\mathfrak{A}_0$ for a sequence of mutually independent subalgebras and $H(\mathfrak{A}_0) = H(T(\mathfrak{A}_0)^0_{-\infty} \mid (\mathfrak{A}_0)^0_{-\infty}) = H(T\gamma_0 \mid \gamma_0)$.

3. Invariants of automorphisms

Every automorphism of the type $\mathfrak{K}_0$ (the subscript is placed in order to distinguish this case from the case of the flows with continuous time) has the Lebesgue spectrum of countably infinite multiplicity, i.e. among the classes $\mathfrak{K}_0^\gamma$ only the class $\mathfrak{K}_0^\omega \subseteq L_0^\omega$ is non-empty. This class splits into subclasses $\mathfrak{K}_0^\omega(h)$ according to the values $h(T)$.

For every h, $0 < h \leqslant \infty$, there are automorphisms with $h = h(T)$.

Examples with $h > 0$ result from the scheme of independent random trials $\dots L_{-1}$, $L_0, L_1, \dots, L_t, \dots$ with the probability distribution ξ_t of the trial L_t given by

$$\mathbf{P}(\xi_t = a_i) = p_i, \qquad -\sum_{i=1}^{\infty} p_i \log p_i = h. \tag{5}$$

The space M is assembled from sequences $x = (\dots, x_{-1}, x_0, x_1, \dots, x_t, \dots)$, $x_t = a_1, a_2, \dots$, and the translation $Tx = x'$ 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 (5).

4. Invariants of flows

In the case of flows we set $h(\{S_t\}) = H(S_1)$. For every h, $0 < h < \infty$, there exist a flow of the class $\mathfrak{K}_0^\omega(h)$, i.e. a flow having Lebesgue spectrum of countably infinite multiplicity and with the prescribed value of the constant h.

The analogy with Section 3 naturally suggests the idea of replacing for the proof of Theorem 4 the discrete independent trials by the "processes with independent increments" or by generalized processes "with independent values" [19, 20]. But such an approach leads only to flows of the class $\mathfrak{K}_0^\omega(\infty)$ [6]. In order to get finite values h one needs to use some more artificial construction. It is possible to present in this note only an outline of one such construction.

Let us define independent random variables ξ_n, corresponding to all integers n, with the following probability distributions of their values: $\mathbf{P}(\xi_0 = k) = 3 \cdot 4^{-k}$, $k = 1, 2, \dots$, and $\mathbf{P}(\xi_n = k) = 2^{-k}$, $k = 1, 2, \dots$, for $n \neq 0$. In the case $\xi_0 = k$, let us consider a random point τ_0 of the t-axis with uniform probability distribution in the interval $-u2^{-k} \leqslant \tau_0 \leqslant 0$, and let us define random points τ_n for $n \neq 0$ by the relation $\tau_{n+1} = \tau_n + u2^{-\xi_n}$.

Let $\varphi(t) = \xi_n$ for $\tau_n \leqslant t < \tau_{n+1}$. It is easy to check that the distribution of the random function $\varphi(t)$ is invariant with respect to the translations $S_t\varphi(t_0) = \varphi(t_0 - t)$. It is easy to calculate that $h\{S_t\} = 6u^{-1}$ (during a unit of time one encounters $3u^{-1}$ points τ_n in average, and every ξ_n contributes the entropy $\sum_{i=1}^{\infty} k2^{-k} = 2$).

One can get a more graphic description of our stochastic process if we include into the description of its state $\omega(t)$ at moment of time t, in addition to the value $\varphi(t)$, the value $\delta(t) = t - \tau^*(t)$ of the difference between t and the closest to t from the left point τ_n. If described in this way, our process turns out to be a stationary Markov process. It deserves to be called only "quasiregular", because, while the corresponding dynamical system is transitive, the value of the difference $f(\omega(t), t) = \tau^*(t) = t - \delta(t)$ is determined up to a dyadic-rational summand by the behaviors of the process realization in any arbitrary far past.

I am grateful to Y. G. Sinai for his participation in the preparation of this paper.

References

[1] Kolmogorov A. N. A new metric invariant of transitive dynamical systems and of automorphisms of Lebesgue spaces (Russian), *Dokl. Akad. Nauk SSSR (N.S.)*, 1958, V. 119, No. 5, p. 861–864. ‡

[2] Rokhlin V. A., Selected topics from the metric theory of dynamical systems (Russian), *Uspehi Matem. Nauk (N.S.)*, Vol. 4 (1949), No. 2(30), 57–128.

[3] Gelfand I. M., Fomin S. V., Geodesic flows on manifolds of constant negative curvature (Russian), *Uspehi Matem. Nauk (N.S.)*, Vol. 7 (1952), No. 1(47), 118–137.

[4] Fomin S. V., On dynamical systems in a space of functions (Russian), *Ukrain. Mat. Žurnal*, Vol. 2, (1950), No. 2, 25–47.

[5] Itô K. Complex multiple Wiener integral, *Jap. J. Math.*, Vol. 22 (1952), 63–86.

[6] Itô K. Spectral type of the shift transformation of differential processes with stationary increments, *Trans. Amer. Math. Soc.*, Vol. 81 (1956), No. 2, 253–263.

[7] Doob J. L. Stochastic processes, *John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London*, 1953, viii+654 pp. Russian translation: *Izdatel'stvo inostrannoi literatury*, 1956.

[8] Shannon C. E., Weaver W. The Mathematical Theory of Communication, *The University of Illinois Press, Urbana, Ill.*, 1949, vi+117 pp.

[9] Gelfand I. M., Kolmogorov A. N., Yaglom A. M., On the general definition of the quantity of information (Russian), *Dokl. Akad. Nauk SSSR (N.S.)*, Vol. 111 (1956), 745–748.

[10] Shannon C. E., A Mathematical Theory of Communication, *The Bell Systems Technical Journal*, Vol. 27 (1948), 379–423, 623–656, *Russian translation.* In: Teoriya

‡An English translation is included into this collection.

peredachi electricheskih signalov pri nalichii pomeh, 7–87, *Izdatel'stvo inostrannoi literatury*, 1953.

[11] Plessner A. I., Functions of maximal operators (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 23 (1939), No. 4, 327–330.

[12] Plessner A. I., Semiunitary operators (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 25 (1939), 708–710.

[13] Sinai Y. G., Dynamical systems with countably infinite Lebesgue spectrum, I (Russian), *Izv. Akad. Nauk SSSR Ser. Mat.* Vol. 25 (1961), 899–924.

[14] Parasyuk O. S., Flows of horocycles on surfaces of constant negative curvature (Russian), *Uspehi Matem. Nauk (N.S.)* Vol. 8, (1953), No. 3(55), 125–126.

[15] Gurevich B. M., The entropy of the horocyclic flow, *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, vol. 136 (1961), No. 4, 768–770.

[16] Kolmogorov A. N., On the entropy per unit of time as a metric invariant of automorphisms, *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, V. 124 (1959), No. 4, 754–755.

[17] Sinai Y. G., On the concept of entropy for a dynamic system (Russian), *C. R. (Doklady) Acad. Sci. URSS (N. S.)*, Vol. 124 (1959), No. 4, 768–771.

[18] Sinai Y. G., On a weak isomorphism of transformations with invariant measure, (Russian) *Mat. Sb. (N.S.)*, Vol. 63(105) (1964), No. 1, 23–42.

[19] Itô K. Stationary random distributions, *Mem. Coll. Sci. Univ. Kyoto. Ser. A. Math.*, Vol. 28 (1954), 209–223.

[20] Gelfand I. M., Generalized random processes, (Russian) *Dokl. Akad. Nauk SSSR (N.S.)*, Vol. 100 (1955), 853–856.

Translator: http://nikolaivivanov.com