The Annals of Probability
1989, Vol. 17, No. 3, 833–839

KOLMOGOROV'S WORK ON ERGODIC THEORY

By Ya. G. Sinai

L. D. Landau Institute of Theoretical Physics

Apparently the interests of Kolmogorov in ergodic theory had already started in the 1930s. In mathematical Moscow it was a period of construction of the foundations of the theory of stationary random processes. One might recall the paper by Khintchine [11] at that time dedicated to the spectral theory of such processes. Slightly later there appeared a very important paper by Kolmogorov [K116][1] on regularity and extrapolation of stationary random sequences. The paper by Khintchine [10], where he gave a purely metric proof of the Birkhoff ergodic theorem, belonged to ergodic theory itself. In view of this paper the ergodic theorem on a.e. convergence of time averages is often called the Birkhoff–Khintchine theorem at least in the Russian literature. In the 1930s, the well-known paper by Krylov and Bogolyubov [12] on invariant measures for groups of homeomorphisms of topological spaces was written.

In the beginning of the 1930s, there appeared the famous paper by von Neumann [21], where the general notion of the metric isomorphism of one-parameter groups of measure-preserving transformations was introduced. Also in [21] von Neumann proved a basic theorem of metric isomorphism of ergodic dynamical systems with pure point spectrum. This theorem showed that for this class of systems the spectrum is the complete metric invariant. Since that time the problem of metric classification of dynamical systems became one of the central ones in ergodic theory.

The scientific activity of von Neumann was always under close attention. It is not surprising that this problem became well known quite soon in Moscow and several mathematicians spent a lot of effort trying to make some progress here. Unfortunately, I have no evidence whether the works of Morse, Hedlund and Hopf on ergodic properties of geodesic flows and on symbolic dynamics were widely discussed at that time.

For Kolmogorov the end of the 1930s was the beginning of his classical works on hydrodynamics and turbulence. His [illegible] publication which can be considered as relating to ergodic theory goes b[illegible] to 1937 (see [K84, K99]), where he exposed the Birkhoff–Khintchine theorem in probabilistic terms. At this time there were contacts between Kolmogorov and Rokhlin who at that time was a student of the mathematics department of Moscow University. For Rokhlin ergodic theory was the main domain of investigations at that time. As he recalled later Kolmogorov understood quite clearly the connections between the theory of stationary random processes and ergodic theory and, in particular, the

Received August 1988.

[1]Reference citations preceded by K refer to the list of Kolmogorov's publications on pages 945–964.

ergodic properties of Gaussian stationary processes. This was much before the appearance of Itô's theory of multiple stochastic integrals which gave the natural approach to these problems. The ideas of Kolmogorov were realized much later in the paper of his student Fomin [5]. Also, Kolmogorov had the idea of the example of the dynamical system with continuous simple spectrum based on the theory of Gaussian stationary processes; the corresponding construction was done explicitly by Girsanov only in 1958 (see [7]).

Kolmogorov was very much excited by the fact that many natural examples of dynamical systems had countable Lebesgue spectrum. It was observed by Halmos and Rokhlin for the case of ergodic group automorphisms of compact abelian groups ([8] and [15]) and by Gel'fand and Fomin [6] for the case of geodesic flows on manifolds of constant negative curvature. The paper by Gel'fand and Fomin [6] was very important not only for ergodic theory but also for the theory of group representations and had many important implications.

From the point of view of probability theory the statement that a dynamical system has countable Lebesgue spectrum means roughly speaking the following. Assume that a stationary random process has a spectral density. If the dynamical system generated by this process has countable Lebesgue spectrum, then any process generated by a nonlinear functional of the original process also has a spectral density. Already in the 1950s before his paper on entropy appeared (see below), Kolmogorov found a general statement which in a sense explains the appearance of countable Lebesgue spectrum. He proved that any measure-preserving mapping generated by a regular stationary sequence has this type of spectrum. This regularity means that the process satisfies the "0–1" law of Kolmogorov. One can find a hint of this result in [K280] which we are going to discuss. The connection of his result with the results of [6], [8] and [15] became clear only later.

In the autumn of 1957, Kolmogorov organized his famous seminar on dynamical systems and gave a lecture course on the same subject. Among the participants and listeners there were Alekseev, Arnol'd, Girsanov, Meshalkin, Pinsker, Sinai, Sitnikov, Tikhomirov and others. Some information about this seminar can be found in [2].

The lectures by Kolmogorov started with the proof of the metric isomorphism of dynamical systems with pure point spectrum. He gave an entirely probabilistic exposition of the corresponding theorem by von Neumann. Later it was found that a similar approach was presented in the book by Blanc-Lapierre and Fortet [3]. In the seminar the participants discussed in much detail the construction of Itô's multiple stochastic integrals and the ergodic properties of Gaussian stationary processes. It is well known that such processes can be obtained as natural limits of quasi-periodic processes, that is, processes corresponding to dynamical systems with pure point spectrum. A general feeling at that time was that there exist some boundary separating dynamical systems of probability theory and dynamical systems appearing in ordinary differential equations, classical mechanics and hydrodynamics or, as we call them sometimes, classical dynamical systems. At the same time Kolmogorov emphasized many times (see his famous lecture at the Amsterdam Congress of Mathematicians [K243]) that such a complicated phenomenon as turbulence should be described by invariant measures for Navier–Stokes equations with good properties of mixing, that is, sufficiently regular random fields.

At the end of 1957, quite unexpectedly for all of us, Kolmogorov wrote his wonderful paper on the entropy of dynamical systems [K280]. Looking back now, it is clear that this paper had in fact many roots. For several years Kolmogorov was involved in problems of information theory. He was probably one of the first mathematicians who realized the importance of Shannon's paper for mathematics. His general talk [K272] had a great influence on mathematicians and engineers. During the All-Union Congress of Mathematicians in 1956, he gave a talk on the joint work with I. M. Gel'fand and A. M. Yaglom on the notion of information and its properties [K267]. Also during these years Kolmogorov and Tikhomirov were working on a large program concerning the estimation of ε-capacities of various compacts in functional spaces (see [K285]) and the corresponding notion of dimension in functional spaces. Kolmogorov intended to connect these questions with the problems of complexity of algorithms and numerical methods.

Returning to his paper on entropy, Kolmogorov first gave, in his lecture, a proof of the theorem which in modern terms is as follows: Any two Bernoulli generators have the same entropy. In the text which was prepared for publication the whole approach and the theorems were quite different. First, Kolmogorov introduced the notion of quasiregular dynamical system or as we say now, of K-systems. (For a short period we used the term "Kolmogorov systems" but he himself asked to replace it by the abbreviated one hoping to have a chance to write a paper on K-systems. As he added, it would be inconvenient for him to write on Kolmogorov systems. Unfortunately this did not happen but the abbreviated term remained). The notion of a K-system is a generalization of the notion of a regular stationary process to ergodic theory. Assume that (Ω, F, P) is a probability space. A group $\{S^t\}$ of measure-preserving transformations, $t \in R^1$ or Z^1 is called a K-system if there exists a σ-subalgebra $F_0 \subset F$ such that

$$F_t = S^t F_0 \supset F_0 \quad \text{for } t > 0, \tag{1}$$

$$\bigvee_t F_t = F, \tag{2}$$

$$\bigwedge_t F_t = N, \tag{3}$$

where N is the trivial σ-algebra of sets of measure 1 and 0.

The subalgebra F_0 is called a K-subalgebra. Kolmogorov defines entropy per unit of time of the dynamical system to be equal to $H(F_1|F_0)$, where $H(F_1|F_0)$ is the conditional entropy. Then he proves that the $H(F_1|F_0)$ does not depend on the choice of the K-subalgebra F_0. It seems now that the idea of introducing K-systems was connected with the idea mentioned above of the difference between dynamical systems in probability theory and classical dynamical systems. Later it turned out that the notion of K-systems became very important in ergodic theory mainly in connection with classical dynamical systems and the theory of deterministic chaos.

In [K280] Kolmogorov also proposed some examples of flows ($t \in R^1$) with finite positive entropy. the positivity was not surprising but finiteness needed some effort. The examples looked rather strange from the probabilistic point of view because they belonged to the class of processes with "discrete interference of randomness." But as was realized several years later precisely this class of processes appears if we try to represent classical dynamical systems as random processes.

After having written the text and submitting it for publication Kolmogorov went to Paris for half a year. At that time I was puzzled by the fact that any measure-preserving group of transformations gives rise to stationary random processes. Based on this, I invented a definition of entropy suitable for an arbitrary dynamical system. However, it was not clear how one could use it and whether it coincided with Kolmogorov's definition. The definition is explained in detail in Shirayev's paper (this issue) and I shall not repeat it here. At that time I met Rokhlin who suggested calculating the entropy of an ergodic group automorphism of the two-dimensional torus. The theorem on generators played a significant role in calculating the entropy.

Generators permit one to represent dynamical systems as stationary processes in a one-to-one correspondence. However, this representation is in no way unique. The theorem on generators states that for the finite generators, that is, for representations by processes with a finite number of states and with $t \in Z^1$ the value of entropy does not depend on the choice of a generator. The proof presented in [18] was very close to Kolmogorov's proof as given in his lecture for the Bernoulli generators. But at one point we have for Bernoulli generators some equality which in the general case is replaced by an inequality. Now it looks fairly trivial but at that time it took some effort to understand that one always has the corresponding inequality.

Studying group automorphisms of the two-dimensional torus, I tried to prove that their entropy is 0 because they belong to classical dynamical systems. I did not succeed and came to Kolmogorov with my drawings. He looked at them and immediately said that the entropy should be positive. After that I got the final result very quickly (see [18]).

Group automorphisms of the two-dimensional torus were the first examples of classical dynamical systems with positive entropy. Afterwards many such systems, such as geodesic flows on manifolds of negative curvature or, more generally, Anosov systems, dispersed billiards, systems with strange attractors, various classes of one-dimensional mappings, and others, were discovered. The general property of dynamics which leads to the positivity of entropy is the instability of the motion. The instability means that trajectories which start at close points diverge exponentially quickly. Entropy characterizes in a sense the speed of this divergence. Now the theory of classical dynamical systems with positive entropy is a well-developed branch of ergodic theory. The connections with equilibrium statistical mechanics are very important here (see [4] and [17]).

The definitions of entropy showed that entropy is a metric invariant of a dynamical system. In other words isomorphic dynamical systems should have the same entropy. Bernoulli shifts or stationary sequences of independent random variables have countable Lebesgue spectrum and may take an arbitrary value of entropy. Thus Kolmogorov had a great breakthrough in the problem of metric isomorphism of dynamical systems showing that two Beroulli shifts with different values of entropy are nonisomorphic. This means that entropy is a new essential invariant independent of the spectrum. Kolmogorov's result showed even more. Now the problem of isomorphism at least for systems with positive entropy looked like a problem of information theory or, more precisely, a problem of stationary coding and not an algebraic problem.

After Kolmogorov's paper the natural problem arose whether two Bernoulli shifts with the same entropy are isomorphic, or, more generally, whether two K-systems with the same value of entropy are isomorphic. The first explicit nontrivial example of isomorphism was constructed by Meshalkin [13]. In [19] it was shown that every ergodic dynamical system has Bernoulli factors with the same value of entropy. The complete solution of the isomorphism problem for Bernoulli shifts was given by Ornstein [14] who invented for this a new, very profound technique. This technique enabled Ornstein and his colleagues to obtain many important results here. In particular it was shown that there exist uncountably many nonisomorphic K-systems with the same entropy.

During the active study of dynamical systems with positive entropy Rokhlin found (see [K284]) that Kolmogorov's approach to the definition of entropy contained an error. Namely, Kolmogorov uses the equality

$$\bigcap_t (A_0 \vee F_t) = A_0.$$

Here A_0 is a K-subalgebra and F_t is a family of subalgebras entering the definition of K-system. Rokhlin gave a concrete example of the A_0, F_t for which the last equality is wrong and the corresponding conditional entropies have different values. Now even simpler examples of K-subalgebras F_0 are known for which $H(F_1|F_0)$ is less than the entropy of dynamical system. But as far as I know it is still an open question whether there exist K-subalgebras with an arbitrary value of $H(F_1|F_0)$ less than the entropy of the whole dynamical system. Kolmogorov wrote another paper where he suggested a third definition of entropy but soon Rokhlin showed [16] that this definition in fact coincides with that given in [18].

Now we shall discuss briefly the papers by Kolmogorov concerning the behavior of Hamiltonian systems close to integrable ones which had a great influence in ergodic theory, dynamical systems and the whole field of nonlinear physics. This part of his activity concerned classical dynamical systems and led to the discovery of the famous KAM theory (Kolmogorov–Arnol'd–Moser theory). The results of Kolmogorov are fully exposed in his talk at the International Congress of Mathematicians in Amsterdam in 1954 [K243]. The text is extremely well written in the original style of Kolmogorov. There he not only explains the results obtained previously, but also foresees the future development of the theory. The paper [K243] had a great influence on many generations of mathematicians and physicists. Returning to the content of KAM theory, it was more or less widely known that the phase space of integrable nondegenerate Hamiltonian systems could be decomposed into invariant tori whose dimension was equal to one-half of the dimension of the whole phase space. On each torus the system has a pure point spectrum. Now this general fact is explained with the help of the so-called Liouville–Arnol'd theorem [1]. But at that time it was mainly a general observation. Like Poincaré, Kolmogorov considered small perturbations of integrable systems and proved that most invariant tori in the measure-theoretic sense are preserved under small perturbations. In fact, he invented a special new perturbation scheme which later became one of the most important and widely used in all of nonlinear dynamics. Later Arnol'd extended it to some important cases of degenerate Hamiltonian systems and Moser, among other results, has shown how to treat Hamiltonians having only a finite number of derivatives. One of the important results of KAM theory is the general statement that nonergodic Hamiltonian systems contain a nonempty open subset. This contradicted claims which one could often see in the physical literature according to which any typical Hamiltonian system with interaction should be ergodic.

Later Kolmogorov put forward several times a hypothesis according to which a typical Hamiltonian system is a mixture of ergodic invariant tori and components of positive measure having positive entropy. As far as I know he did not work seriously on this problem and did not make any concrete suggestions.

Kolmogorov had a great influence on many mathematicians of various generations proposing very interesting problems and giving some ideas of how to approach the solutions. For example, the famous problem concerning the asymptotic behavior of invariant measures of diffusion processes when the diffusion matrix goes to 0 was already proposed by Kolmogorov in the 1950s. The first results here were obtained by Hasminski [9]. Later the problem was considered in larger generality by Ventzel and Freidlin (see monograph [20]), Kifer and others. Kolmogorov made very profound remarks concerning statistical physics and phase transitions. For all Kolmogorov's students it was a great pleasure to collaborate with him, to work in directions of which he approved and to admire his great personality. These years for each of us are unforgettable.

Acknowledgments. I thank Sh. Newhouse and C. Series for their help with the manuscript.

REFERENCES

[1] Arnol'd, V. I. (1978). *Mathematical Methods of Classical Mechanics*. Springer, New York. (Original Russian ed., Moscow, 1974.)

[2] Arnol'd, V. I. and Meshalkin, L. D. (1960). Kolmogorov's seminar on selected problems of analysis (1958–1959). *Uspekhi Mat. Nauk* **15**(1) 247–250. (In Russian.)

[3] Blanc-Lapierre, A. and Fortet, R. (1953). *Theorie des fonctions aléatoires*. Masson, Paris.

[4] *Dynamical Systems-2. Ser.* (1985). *Modern Problems of Mathematics. Fundamental Directions*. VINITI, Moscow.

[5] FOMIN, S. V. (1950). On dynamical systems in functional spaces. *Ukrain. Mat. Zh.* **2**(2) 25–47. (In Russian.)

[6] GEL'FAND, I. M. and FOMIN, S. V. (1952). Geodesic flows on manifolds of constant negative curvature. *Uspekhi Mat. Nauk* **7**(1) 118–137. (In Russian.)

[7] GIRSANOV, I. V. (1958). On spectra of dynamical systems generated by Gaussian stationary processes. *Dokl. Akad. Nauk SSSR* **119** 851–853. (In Russian.)

[8] HALMOS, P. R. (1943). On automorphisms of compact groups. *Bull. Amer. Math. Soc.* **49** 619–624.

[9] HASMINSKI, R. Z. (1963). Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion. *Theory Probab. Appl.* **8** 1–21.

[10] KHINTCHINE, A. YA. (1933). Zur Birkhoff's Lösung des Ergodenproblems. *Math. Ann.* **107** 485–488.

[11] KHINTCHINE, A. YA. (1938). The correlation theory of stationary stochastic processes. *Uspekhi Mat. Nauk* **5** 42–51. (In Russian.)

[12] KRYLOV, N. M. and BOGOLYUBOV, N. N. (1937). La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Ann. of Math.* (2) **38** 65–113.

[13] MESHALKIN, L. D. (1958). One case of isomorphism of Bernoulli schemes. *Dokl. Akad. Nauk SSSR* **128** 41–44. (In Russian.)

[14] ORNSTEIN, D. (1974). *Ergodic Theory, Randomness and Dynamical Systems.* Yale Univ. Press, New Haven, Conn.

[15] ROKHLIN, V. A. (1949). On endomorphisms of compact abelian groups. *Izv. Akad. Nauk SSSR Ser. Mat.* **13** 323–340. (In Russian.)

[16] ROKHLIN, V. A. (1963 / 65). Generators in ergodic theory. I, II. *Vestnik Leningrad. Univ.* **18**(1) 26–32, **20**(13) 68–72. (In Russian.)

[17] RUELLE, D. (1978). *Thermodynamic Formalism.* Addison-Wesley, Reading, Mass.

[18] SINAI, YA. G. (1959). On the concept of entropy for dynamic system. *Dokl. Akad. Nauk SSSR* **124** 768–771. (In Russian.)

[19] SINAI, YA. G. (1964). Weak isomorphism of measure-preserving transformations. *Mat. Sb.* **63** 23–42. (In Russian.)

[20] VENTZEL, A. D. and FREIDLIN, M. I. (1979). *Fluctuations in Dynamical Systems under Small Random Perturbations.* Nauka, Moscow. (In Russian.)

[21] VON NEUMANN, J. (1932). Zur Operatorenmethode in der klassischen Mechanik. *Ann. of Math.* (2) **33** 587–642.

L. D. LANDAU INSTITUTE OF THEORETICAL PHYSICS
U.S.S.R. ACADEMY OF SCIENCES
GSP-1 117940 KOSYGIN ST. 2
MOSCOW V-334
U.S.S.R.