OBITUARY

ANDREI NIKOLAEVICH KOLMOGOROV
(1903–1987)

A tribute to his memory organised by David Kendall

CONTENTS

ACKNOWLEDGEMENTS. We wish to express our gratitude to those who have helped us in the production of this work. Professor A. N. Shiryaev in his capacity as Kolmogorov's literary executor kindly provided us with an extensive bibliography (here further enlarged and translated into English by E. J. F. Primrose, who also wrote the commentary on it). Professor Shiryaev also gave us the three photographs, and allowed DGK to have access to his definitive biography. Professor Yu. A. Rozanov added several illuminating details to DGK's introductory essay.

DGK acknowledges support from the Science and Engineering Research Council.

KOLMOGOROV: THE MAN AND HIS WORK

D. G. KENDALL

The subject of this memoir (ANK in what follows) was born on 25 April 1903 in Tambov during a journey from the Crimea to his mother's home. He was the son of parents not formally married; ANK's mother Mariya Yakovlevna Kolmogorova was one of three sisters who have been described as independent women with lofty social ideas. Mariya Yakovlevna died in childbirth at Tambov and her son Andrei Nikolaevich was adopted and brought up in the village of Tunoshna (near to Yaroslavl on the river Volga) by her sister Vera Yakovlevna Kolmogorova. To her nephew Vera Yakovlevna gave all the love of a mother, and Andrei Nikolaevich responded with all the love of a son. It is warming to record that Vera Yakovlevna lived until 1950, so that she was to witness some of ANK's greatest achievements.

Andrei Nikolaevich is always known to us by the family name of his maternal grandfather Yakov Stepanovich Kolmogorov, a leading member of the Uglich nobility. It was in the Kolmogorov home at Tunoshna that ANK spent his earliest years. During his childhood this was a centre of clandestine printing, and family traditions record that on occasion compromising documents were hidden under his cradle.

Of ANK's father, Nikolai Kataev (I have not been able to find his patronymic), we know that he was the son of a priest, that he became a professionally trained agriculturalist, that he was exiled to Yaroslavl, that after the Revolution he became a department head in the Agriculture Ministry, and that he perished on the southern front during the offensive by Denikin in 1919.

In 1920 ANK went to Moscow University as a student of mathematics, but also attended lectures in metallurgy. In addition to this he participated in Bakhrushin's seminar on Russian history, where he presented the results of his first piece of research – on landholding in Novgorod in the 15th–16th century. We are told how this was received by his teacher: 'You have supplied one proof of your thesis, and in the mathematics that you study this would perhaps suffice, but we historians prefer to have at least ten proofs.' This anecdote is usually told as a joke, but to those who know something of the limitations of such archives it will seem a fair comment. However that may be, it is on record that an expedition organised by P. S. Kuznetsov later confirmed ANK's conjecture about the way in which the upper Pinega was settled.

Aleksandrov, Luzin, Suslin, and Uryson all helped in various ways to stimulate ANK's early mathematical researches, but it appears that his principal teacher at that time was Stepanov. In 1922 he produced a synthesis of the French and Russian work on the descriptive theory of sets of points. This was not published until 1928, and then only in part [1928a]. The full text eventually appeared in his collected works. At about the same time he was introduced to Fourier series (in Stepanov's seminar) and in 1922 he discovered that there is no slowest rate of convergence to zero for the Fourier cosine coefficients of a summable function [1923b]. In 1963 I gave a lecture in Tbilisi in which I showed that in the transient aperiodic case the diagonal Markov transition probabilities $p_{ii}^{(n)}$ form a sequence of Fourier cosine coefficients, and remarked that it would be interesting to see what could be deduced from this about their rate of convergence to zero as n tends to infinity. ANK offered the comment that 'Il existe une possibilité purement analytique...', the gist of which escaped me at the time because of language difficulties. It occurs to me now that he may have been thinking of an application of his 1923 paper in this new context. To the best of my knowledge that comment has not yet been followed up.

Plate I shows ANK (wearing spectacles, leaning over to his left) at the Tbilisi conference. Also in the picture are Dynkin and Gnedenko (to ANK's right) and many other celebrated probabilists.

In 1922 ANK also constructed his famous example of a summable function whose Fourier series diverges almost everywhere [1923a]. ANK was 19 years old at the time, but suddenly he had become an international celebrity, the more so after he had sharpened this result from 'almost everywhere' to 'everywhere' [1925h]. For more details of this and of ANK's other principal achievements the reader must turn to the accompanying articles by other writers. Here I will in the main chronicle only ANK's life, but my account will be seasoned with remarks about a few of his discoveries and their implications that do not fit so well into the scope of any of the more specialised contributions. Let us note, however, one highly significant date: in 1925 there appeared a paper [1925f], written jointly with Khinchin, that represents ANK's first involvement with probability theory. It contains a proof of the 'three series' theorem, as well as the Kolmogorov inequality involving the maxima of partial sums of independent random variables (whence, ultimately, the martingale inequalities and the stochastic calculus.)

ANK became a postgraduate student in 1925, supervised by Luzin. He emerged from the postgraduate school of 1929 with 18 mathematical papers to his credit. These included his versions of the strong law of large numbers and the law of the iterated logarithm, some generalisations of the operations of differentiation and integration, and a contribution to intuitionistic logic. His two papers [1925e, 1932d] on this last topic are regarded with awe by specialists in the field. The Russian language edition of Kolmogorov's collected works contains a retrospective commentary on these papers which ANK evidently regarded as marking an important development in his philosophical outlook.

From the summer of 1929 dates his lifelong friendship with Aleksandrov, which started with a 21-day trip starting from Yaroslavl, first by boat down the Volga, then on to Samara, the Caucasus, and Lake Sevan in Armenia. On the shores of the lake Aleksandrov worked on chapters of his joint book with Hopf, while ANK brooded over what was to be his paper [1931a] on Markov processes with continuous states in continuous time. Modern diffusion theory dates from that work, though it is analytical, and sample paths do not appear in it. We note in passing that path-theoretic diffusion was to grow out of the earlier work by Bachelier and by Wiener, to whom we owe the basic mathematics of brownian motion. (It is always relevant to add that Wiener was directed to the study of brownian motion by Bertrand Russell. Wiener had come to Cambridge with the intention of studying logic; Russell discouraged this, and told him to read Einstein's 1905 note **[11]** instead. While Einstein predicted and described Brownian motion in quantitative detail, its precise relationship to the observations made in 1827 by Robert Brown **[5]** was not at first clear. Concerning this Einstein wrote 'Es ist möglich, daß die hier zu behandelnden Bewegungen mit der sogenannten "Brownschen Molekularbewegung" identisch sind; die mir erreichbaren Angaben über letztere sind jedoch so ungenau'.)

What was perhaps most significant about the 1931 paper by ANK was the link with the theory of linear partial differential equations – here he was almost certainly much influenced by Petrovskii. At the time this was a startling development. Today the theory of parabolic and elliptic linear partial differential equations has been merged with the theory of Markov processes, with each discipline lending strength to the other.

A little before this ANK had published his first attempt [1929b] at a foundational paper on probability itself. This was based on measure theory, and introduced elementary events, random events as measurable sets of elementary events, and random variables as measurable functions. There were no sigma-algebras, no conditional expectations, and no stochastic processes.

These omissions were to be filled by ANK's monograph [1933b] entitled *Grundbegriffe der Wahrscheinlichkeitsrechnung*. This was written in the forest on the banks of a small river, and was published in Berlin. In the Foreword he remarks that his aim is to create an axiomatic foundation for probability theory, and he comments that without Lebesgue measure and integration this task would have been hopeless. He also stresses the necessity first to strip away from the Lebesgue theory those elements that tie it too strictly to geometry, and acknowledges the role that Fréchet played in bringing this about. He directs the reader to three particularly novel developments presented in the book: the treatment of probability distributions in infinite-dimensional spaces, the introduction of rules for integrating or differentiating 'under the expectation sign', and the construction of a mathematical theory of conditional probabilities and expectations. The first and third of these topics are methodologically closely allied, as was later made fully explicit by Ionescu Tulcea **[20]**. ANK was careful to stress that the vital tool in the theory of conditioning is the generalisation by Nikodym **[28]** of an earlier theorem in a more classical setting due to Radon. Another treatment of infinite-dimensional distribution theory is to be found in P. J. Daniell **[6, 7]**, but his papers seem to have attracted little notice until much later. The historically important point is that the proper development of stochastic process theory had to wait for a general treatment of conditioning. This Kolmogorov gave, and it would have been impossible without the Nikodym theorem.

The conditioning here is with respect to a σ-algebra defined in terms of a family of conditioning random variables. Filtrations of σ-algebras necessarily occur implicitly in the treatment of infinite-dimensional probability spaces, but it does not appear that the purely *information-theoretic* view of random events, random variables, and σ-algebras had yet surfaced. That was to happen later, and it became fully explicit, together with the related concept of separability, in Doob's book **[10]**.

I should like to record a remark made by ANK during the Amsterdam International Congress of Mathematicians in 1954. A lunch for probabilists had been organised in the Amsterdam Zoo by Jerzy Neyman, and a few apprentices like Harry Reuter and myself were invited to represent the younger generation. During the meal Kolmogorov leaned over and said to Doob 'The whole of the theory of stochastic processes will now be based on your work.' I enjoyed watching Doob's pleasure, unsuccessfully concealed by embarrassment.

Some other anecdotes concerning ANK's respect for other mathematicians can perhaps best be related here. I have already referred to ANK's admiration for Fréchet. Fréchet himself said to me 'How curious it is; Lévy's principal colleague among the Russian probabilists is Khinchin, whereas for me it is Kolmogorov – we once spent a vacation together on the Mediterranean coast.' Years later I referred to this when talking with ANK, and he said at once, 'Pas exactement un collègue, plutôt mon maître.'

On another occasion, in 1967, Hermann Dinges and I organised an Oberwolfach meeting on the analytical theory of branching processes. We invited ANK, and to our delight he accepted and brought several other Soviet mathematicians with him. At first ANK said he just wanted to be a listener, but after several highly theoretical talks he looked rather uncomfortable, and eventually told us that he would after all give a lecture that would perhaps remind people of the biological background to the subject. Inevitably he referred to *The genetical theory of natural selection* **[14]** – 'das wundervolle Buch von R. A. Fisher.' Two United States mathematicians sitting near to me were overheard to whisper 'It *can't* be the R. A. Fisher *we* know'.

There is another half to that story. Will Feller used to say that if Kolmogorov had not written his 1931 paper, the whole of stochastic diffusion theory would eventually have been pieced together starting with the ideas in Fisher's book.

But to return to the *Grundbegriffe*: ANK there illustrates his second 'new feature' by an application of it to geometrical probability, and refers to the interesting joint paper with Leontovich [1933d] for a more detailed example. This is perhaps the first occurrence in the literature of a careful treatment of the expectation of the measure of a random set.

The 'backwards' and 'forwards' partial differential equations in the 1931 paper can be thought of as differentiated versions of what is called the Chapman–Kolmogorov equation encapsulating the semigroup property inherent in all markovian situations. I once asked Sydney Chapman about 'Chapman–Kolmogorov', and was surprised to find that he did not know of that terminology. Of course physicists have their own names for such equations. The original Chapman reference seems to be to his 1928 paper in volume 119 of Series A of the Royal Society's *Proceedings*: 'On the brownian displacements and thermal diffusion of grains suspended in a non-uniform fluid'.

In 1931 ANK became a Professor in Moscow University. Just before this he and Aleksandrov made a long scientific trip through Germany (Berlin, Göttingen and Munich) and France (Paris, and the Mediterranean). In Paris there were long talks with Lévy, and a month by the sea was spent with Fréchet, who later told me that they stayed in a lodging house whose lady proprietor had recently installed modern plumbing – unique in that township. Thus they found themselves invited to a party for the whole community, to celebrate – with champagne – the first flush.

In the nineteen-thirties ANK's work started to ramify. What we think of as classical probability theory still occupied much of his time – this was the period during which the stable laws, the infinitely divisible laws, and the theories related to these were being studied by a now growing school of colleagues and pupils, but it also saw ANK's independent development of cohomology theory [1936f], his necessary and sufficient condition for a topological vector space to be normable [1934b], his numerous contributions to approximation theory, his contribution [1933e] to the Kolmogorov–Smirnov test that is expressed in terms of the empirical distribution function, his theory of the structure and limiting behaviour of homogeneous countable Markov chains [1936g, 1936i], his theory of statistical reversibility [1937g], his introduction [1935c] of the characteristic functional (with an eye to applications in non-linear quantum mechanics), his inequalities [1938o] for the moduli of high derivatives (linked to the theory of quasi-analytic functions), his work with Gel'fand on rings of continuous functions defined over topological spaces [1939d], and much else.

The 'much else' included contributions to queuing theory [1931b], to the branching processes of Bienaymé, Watson and Galton and their generalisations [1936q, 1938n], and to the stochastic geometry of the crystallisation process (and of the growth of vegetation) [1937e], as well as (with Piskunov and Petrovskii) the analysis of the solitary waves associated with the spreading of the range of an advantageous gene in a linear habitat [1937d]. This last was written within a year of a similar but independent study by Fisher [**15**]. Both recognised that the range inhabited by the favoured individuals would expand with an asymptotically constant velocity, but the Soviet writers showed that in fact there is a half-infinite interval of possible speeds with each of which there is associated a corresponding travelling wave. This is now a subject in its own right called 'reaction-diffusion theory' (see Britton [**4**]). The application to genetics with which the subject originated has since been joined by applications to the spreading of epidemics [**21**], to the spreading of cultural innovations, to the dynamics of advertising, to the spreading of rumours, and to numerous physical, chemical, and other biological problems.

Another involvement of ANK with work in genetics is of some general interest.

It came about just after he was elected to the Academy of Sciences of the USSR in 1939 (shortly before this he had been appointed Head of the Probability Section in the Steklov Institute). N. I. Ermolaeva **[12]** had reported the results of a botanical experiment to test the claim that in a simple Mendelian situation the proportion r/n of plants displaying the dominant phenotype would have an average value $\frac{3}{4}$. Apparently she had suggested that there were some discrepancies. ANK decided to re-examine her data and to make use of the fact that Mendelian theory predicts not only $\mathscr{E}(r/n)$ but also $\operatorname{var}(r/n)$ and indeed $\mathscr{L}(r/n)$. He therefore plotted the empirical distribution for

$$\Delta = \left(\frac{r}{n}-\frac{3}{4}\right)\frac{4\sqrt{n}}{\sqrt{3}}$$

using her observed numbers, and obtained a good fit to the cumulative distribution for $\mathscr{N}(0;1)$. ANK then wrote: 'This material, despite Ermolaeva's claims to the contrary, has proved to be a new brilliant confirmation of Mendel's laws' [1940g]. And indeed, from ANK's results it is clear that Ermolaeva's experiments must have been carried out with scrupulous care. ANK's uncharacteristic use of the word 'brilliant' makes it plain that he intended his remarks to be taken as an exceptional compliment.

But others did not think so. T. D. Lysenko **[26]** wrote 'in this controversy between Kolmogorov, Member of the Academy, and postgraduate Ermolaeva, it is Ermolaeva who is in the right, and not Kolmogorov'. Lysenko's brief note was followed by a much longer communication from E. Kolman [**24**, communicated by Lysenko]. This should be read in its entirety, for these few quotations may give a wrong impression. If we omit an argument linking Kolmogorov with von Mises, and von Mises with Mach, so that his 'views on the relation of theory to reality are the same as those which were subjected to destructive criticism by Lenin', the key sentence is: 'Now while incompatibility of some material with a given theory disproves the latter, compatibility neither proves nor confirms this theory, for the same material may prove to be compatible also with other theories.' It will be recalled that ANK just claimed '*confirmation*'. Kolman explicitly draws a distinction between proof and confirmation, but seems to rule out the possibility of ever 'confirming' anything. This however was in 1940, and the controversy seems to have petered out as the protagonists, like others elsewhere, began to find themselves confronted with very different tasks. The whole incident is rightly viewed as one throwing light on ANK's strong personality, and his determined pursuit of truth whatever the obstacles.

Two other novel topics interested ANK in the years before the war. The first was the stochastic growth of the area swept out by a circle of fixed radius when its centre follows a two-dimensional brownian motion. This was the joint study with Leontovich [1933d] already mentioned above. The second [1934j] concerned Markov processes describing a random system 'with inertia' whose state at time t is described by all the random variables

$$(q_1, q_2, \dots, q_n, \dot{q}_1, \dot{q}_2, \dots, \dot{q}_n),$$

that is, by the rates of change of the coordinates as well as by the coordinates themselves. This can be seen as a broad generalisation of the Ornstein–Uhlenbeck process.

ANK's immensely influential work on the smoothing and *prediction* of stochastic processes with stationary ordinates (or increments) started as early as 1938 with his paper [1939e] written against a background provided by Khinchin and Slutskii. The topics proved later to be of great military importance, and so it is scarcely surprising that another attack on the problem was mounted by Norbert Wiener in the USA. (One should also mention distantly related work by Harald Cramér in Sweden and by Michel Loève in France.) Of course all such investigations were eventually to be covered by a cloak of secrecy during World War II, but four more of ANK's papers [1940b, c, 1941a, b] were published openly in 1940–41.

In 1947 ANK delivered a comprehensive report on these topics to the General Assembly of the Academy of Science, while in 1949 Wiener in the book *Extrapolation, interpolation, and smoothing of stationary time series* **[31]** expounded his own secret wartime researches in the same area. That book contains a footnote (on page 59) giving Wiener's own view of the historical relationship between the two investigations, and concluding with Wiener's interesting comment that 'the parallelism between ⟨our approaches⟩ may be attributed to the simple fact that the theory of the stochastic process had advanced to the point where the study of the prediction problem was the next thing on the agenda'.

In fact the approaches by Kolmogorov and Wiener complement one another in an interesting way (see the article by Whittle for further detail on this). Insofar as priority in such a confused situation is important, there seems to be no doubt that it was ANK who was first in the field.

In the UK the study of time-series had long been dominated by the periodogram, and the use of this had become unpopular because of its tendency to indicate (as was first thought) a bewilderingly large number of 'periods'. This deflected attention away from the spectrum to the correlation function, until P. J. Daniell in a contribution to a Royal Statistical Society discussion **[8]** pointed out that this was simply the response to the presence of a continuous spectrum, and that the spectrum itself could be satisfactorily estimated by a smoothing procedure applied to the periodogram. A suitable smoothing device was then introduced by M. S. Bartlett **[2]**, and the serious study of such problems in the UK seems to have begun at that point.

The work discussed in the last few paragraphs eventually brought about a profound change in the relationships between probabilists and statisticians on the one hand, and physicists and engineers on the other. No longer could statistics be described (or dismissed) as 'the arithmetic of the social sciences'. Indeed a whole new branch of engineering technology had been created as it were overnight, and now affects almost every aspect of our lives.

From stationary stochastic processes to stationary stochastic fields and thence to the study of turbulence is a natural progression. ANK's interest in turbulence dated from the late thirties, and it was to lead to some of his greatest discoveries. In 1940 he wrote a famous paper [1941d] on the local structure of turbulence, and this was later supplemented by his 'two-thirds' law. He remained much concerned with this subject over a long period. In 1946 he became Head of the Turbulence Laboratory in the Academy Institute of Theoretical Geophysics (to be succeeded in 1949 by his pupil Obukhov), and in 1970–2 he sailed with the scientific research ship *Dmitrii Mendeleev* as Scientific Supervisor of a study of oceanic turbulence. (Some readers will recall a similar episode in the life of Halley.)

I am told by Professor G. I. Barenblatt (formerly one of ANK's students) that there was a substantial interaction between the Cambridge school of fluid dynamicists and Kolmogorov's own group long before there was any possibility of direct contact. He says 'Kolmogorov's two papers on locally isotropic turbulence may now seem to be absolutely transparent, but in the late forties even his own students found them difficult to comprehend. Accordingly G. K. Batchelor's paper on Kolmogorov's theory of locally isotropic turbulence (*Proc. Cambridge Philos. Soc.* 43 (1947) 533–559) came to play an extremely important role in disseminating Kolmogorov's ideas, not only in the West, but also in the USSR itself, and among ANK's closest associates. This happened despite the fact that ANK's students were unable to read English – they used a Russian translation of GKB's paper that has been carefully treasured ever since.'

Before leaving the war period one must mention some of ANK's contributions to specifically military topics. Thus he wrote on theoretical aspects of the effectiveness of fire-control systems, and on the advantages of artificially induced dispersion. It would be interesting, if permitted, to compare his work [1945a,b] with comparable developments made elsewhere.

On the personal side a very important event in his life was his marriage in 1942 to Anna Dmitrievna Egorova.

In the immediate post-war period we find ANK writing on mathematical geology, on inferential statistics (unbiased estimates, quality control), on branching processes (again, now with several types of particle and with a much broadened field of application), and contributing 88 articles to the Soviet Encyclopaedia. It has been suggested that these articles should be translated into English and published together, and certainly this would give us a unique insight into his mathematical thinking that would be extremely valuable. At the same time he was working with B. V. Gnedenko on their book *The limit distributions for the sums of independent random variables* [1949a], which immediately became a classic.

Most of Kolmogorov's papers on probability theory announced and proved major theorems that immediately took their place as foundation stones of the subject, but [1951a] was quite different; it revealed bizarre phenomena (originally called 'pathological') and asked for their investigation. This paper appeared at about the same time as a similar complementary one by Lévy [**25**], and the two works together have had a great influence. Kolmogorov proved that at $t = 0$ the right-handed 1st-order derivatives q_{ij} of a standard Markov transition function $p_{ij}(t)$ always exist (finitely if $i \neq j$), and he constructed an example now called K1 to show that a diagonal element, q_{11} say, of this q-matrix can be equal to $-\infty$. He also proved that

$$\sum_{\alpha \neq i} q_{i\alpha} \leqslant -q_{ii} \text{ for all } i,$$

and in a second example now called K2 he showed that this inequality sign can be strict. He then asked: do the derivatives of *all* orders exist for *all* $t > 0$? To this question Ornstein [**29**] (extending earlier work by Austin) gave an affirmative answer for the first derivative, but Yushkevich [**32**] showed by an example that the answer can be negative for the derivatives of the second and higher orders.

ANK's examples K1 and K2 were studied by Kendall and Reuter [**23**] using semigroup methods, and were later supplemented by others (including a particularly elegant one given in Blackwell [**3**] for which $q_{ii} = -\infty$ for all i), and gradually it became apparent that what was being called pathological behaviour was, in a sense, the norm. Out of this grew a general theory of the sample-path behaviour of such countable-state continuous-time Markov processes developed by Chung, Dobrushin, Doob, Feller, Hou, Neveu, Reuter, D. Williams, and others. One major question in this area, posed in [**22**], remains open, however.

Suppose that $p_{ij}(t)$ and $p^*_{ij}(t)$ are two standard Markov transition functions and that for all i and j there exist numbers $\tau_{ij} > 0$ such that

$$p_{ij}(t) = p^*_{ij}(t) \text{ whenever } 0 \leqslant t \leqslant \tau_{ij}.$$

Does it follow that $p_{ij}(t) = p^*_{ij}(t)$ for all i, j, and t? Recently affirmative answers have been given in a few very special cases by a group of Chinese scholars (Di **[9]**, Hou **[19]**; see also Reuter **[30]**), but the general problem remains open. It seems that an approach using non-standard analysis might be fruitful, but a serious exploration of that possibility has yet to be made.

From the 1950s onward ANK's most important scientific work revolved around the quartet of ideas: probability, dynamics, information, complexity. Some specialist surveys of this area accompany the present memoir, and so I will not go into full detail here, but I will try to comment on at least some of the philosophical implications. It is important to stress that these investigations were indissolubly linked with ANK's profound contributions to mechanics, so that his enquiries were at one and the same time concerned with how we perceive our environment, and how that environment works. The work in mechanics started with his first contribution [1954b] to what has become known as KAM theory (A for Arnold, M for Moser). This was presented at the International Mathematical Congress in Amsterdam in 1954, and there created such a sensation that he was invited to deliver the lecture twice, once in the French and once in the German language (the text published in the ICM Proceedings is in Russian).

Another component of the massive programme just mentioned is ANK's work on the 13th Hilbert problem. Incidentally it is often overlooked that ANK had already solved another of Hilbert's problems (one that asked for a mathematical formulation of probability theory). As we shall see later, there is a sense in which he solved *that* problem *twice*.

The 13th problem was concerned with the possibility of representing a real *continuous* function of many variables by *finite* superpositions of *continuous* functions of fewer variables – one might call it the generalised slide-rule problem. (Actually Hilbert seems to have had in mind the rather different problem in which the functions concerned are supposed to be analytic, or algebraic. The results are then rather different. ANK worked on this 'strict' version of the 13th problem also, and the reader will find more details in the accompanying article by Lorentz.) ANK began his attack on the 13th problem by establishing a special seminar for its study, and later remarked that he had done so without much hope of making progress. But in 1956 [1956f; see also 1955a, 1956d] he was able to announce that each continuous function of any number of real variables can be represented by a finite superposition of continuous functions of only three variables. In 1957 Arnold matched this by showing that 3-variable continuous functions could always be constructed by superposing continuous functions of two variables. (Together these two results settled the 'continuous' case of the original Hilbert problem.) Later in the same year [1957b] ANK showed that every continuous function $h(x_1, x_2, \dots, x_n)$ of a finite number n of variables can be represented in the form

$$\sum_{q=1}^{2n+1} f_q\left(\sum_{i=1}^{n} g_{qi}(x_i)\right),$$

where each f and g is continuous and where the collection of 'inner' functions g_{qt} can be fixed once and for all. ANK is on record as saying that this was technically the most difficult of all his achievements. Essentially it says that 'generalised slide-rules suffice'. Of course careful note must be taken of the fact that continuous functions form a wide class, and can have horrendous properties from a practical point of view (compare brownian paths).

Perhaps as a foil to his information-theoretic programme ANK was at the same time pursuing statistical studies in philology, linguistics, prose- and verse-style, and speech. This work of ANK is little known or appreciated outside the USSR, but was to him a serious and important part of his work in applied mathematics. I remember hearing him give a lecture on 'Applications of probability theory and mathematical statistics to poetics' at the Tbilisi meeting in 1963. The poetry was that of Pushkin, and so it is interesting to recall here Markov's work in 1913 on the sequence of vowels and consonants in a poem by Pushkin in connexion with his development of what we now call Markov chains.

A book on poetic style containing essays by Kolmogorov, the younger Prokhorov, and many others has been published [**17**], and I tried to interest a UK publisher in the possibility of an English translation. The representative of the august University Press in question replied that the book would only interest professional students of Russian literature, and that for them a translation would be unnecessary. As spoken and written language is one of the most important things that we all have in common, and as the joint study of what we hold in common is one of the most effective forms of cooperation, I felt that this lack of interest was a very depressing indicator of the current human condition, and hope that some enterprising publisher will now take up the challenge and commission an English edition of ANK's works [1962b, 1963e, 1963f, 1963g, 1964a, 1964b, 1965f, 1968j, 1968k, 1984c, 1985a] in this area. (The list just given is believed to be complete.)

When one re-reads the *Grundbegriffe* of 1933 with proper attention to the footnotes one is much struck by what Kolmogorov does not say. (Students of the Sherlock Holmes archive will at once recall the curious incident of the dog in the night-time.) He makes many interesting comments, but seems to shy away from any detailed explanation of the relationship between his axioms and empirical practice, referring the reader to the book by von Mises [**27**] (published in 1931) for this.

It is worth noticing that another recently published book by Hostinský [**18**], and the then forthcoming book by Fréchet [**16**] also occur among his references. Two years before this ANK had spent a month with Fréchet shortly after the appearance of the Hostinský and von Mises books, and it is unthinkable that they did not discuss them, especially as Hostinský had there presented Poincaré's 'explication du hasard' in terms of the inevitably discrete (if fantastically fine) partitioning of dynamical phase spaces by a human observer. Years later ANK was to make essential use of that idea in his information-theoretic approach to mechanics.

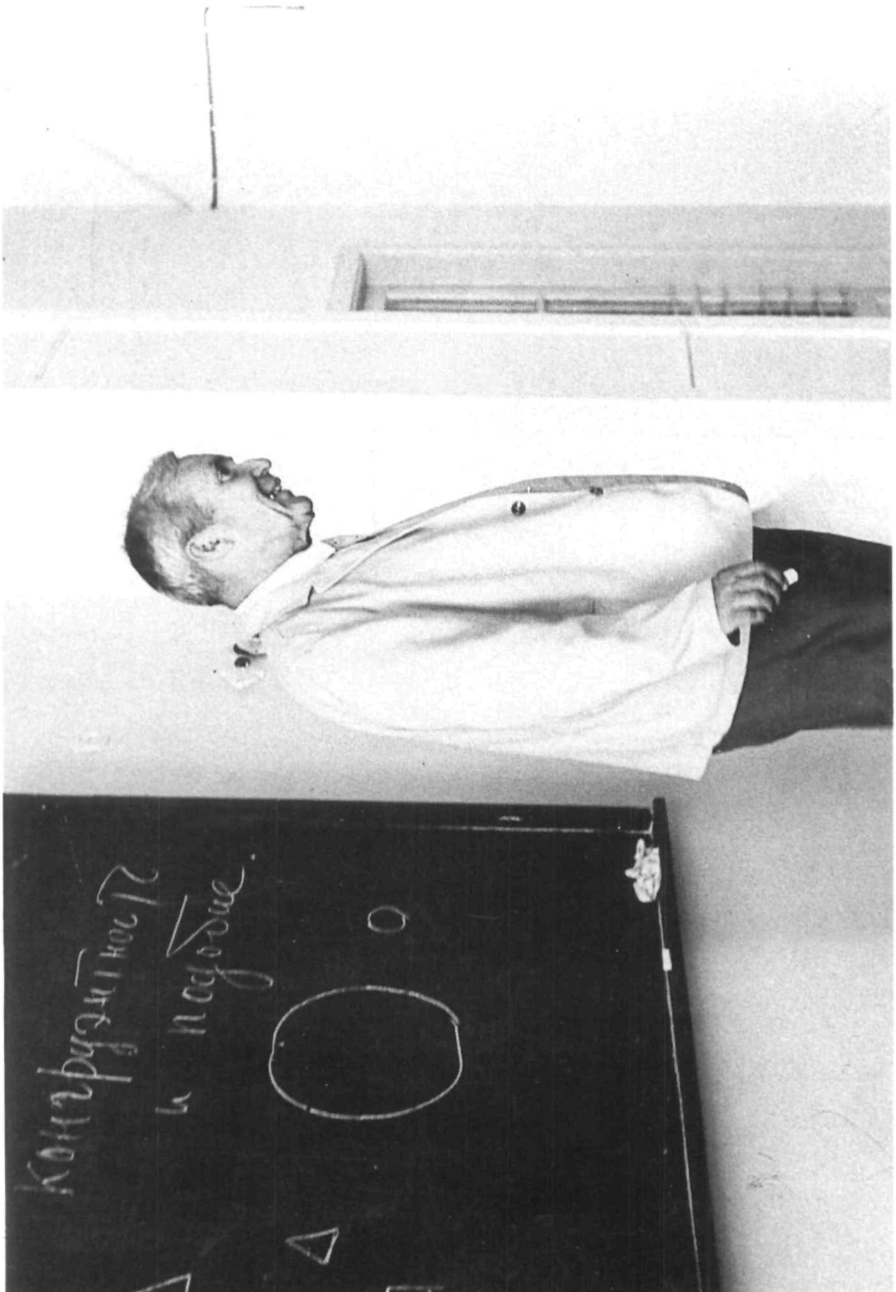
I do not propose to discuss Kolmogorov's work on information theory, ergodic theory, and mechanics because the reader will find expositions of this in the accompanying articles by Razborov, Parry, and Moffatt. But I do want to give just a sketch of ANK's re-formulation of both information theory and probability theory in terms of *complexity*, because that re-formulation is indeed almost a cultural revolution, turning both subjects inside out, and reversing the order in which they are normally considered.

It will, I feel sure, be understood that my account is no more than an outline,

PLATE I

The Tbilisi conference, 1963. See text, p. 40

PLATE II



A. N. Kolmogorov lecturing to his Moscow schoolchildren

Kolmogorov walking in the Caucasus

limited by the slightness of my knowledge and by the intrinsic difficulty of the subject. But there are philosophical (and indeed practical) aspects of it with which we must all become familiar, because it is already clear that the new point of view is likely to percolate throughout the whole of science. For a detailed account the reader is referred to the comprehensive and eloquent exposition by Kolmogorov and Uspenskii [1987c], ANK's own contribution to which may come to be regarded as his scientific testament.

In order not to overburden this presentation I will omit detailed references, but it must not be supposed that ANK carried through this massive programme on his own. On the contrary, vitally important contributions were made by a number of eminent mathematicians including G. Chaitin, A. Shen, R. J. Solomonoff, V. V. Vjugin, and A. K. Zvonkin, as well as those explicitly mentioned below.

The theory is based on a consideration of *finite* objects and *finite* algorithmic operations thereon. The spirit of the programme is summed up in the following quotation from ANK's article [1963d] in the Indian statistical journal *Sankhyā*.

> I have already expressed the view that the basis for the applicability of the results of the mathematical theory of probability to real random phenomena must depend on some form of the frequency concept of probability, the unavoidable nature of which has been established by von Mises in a spirited manner. However, for a long time I had the following views.
>
> (1) The frequency concept based on the notion of limiting frequency as the number of trials increases to infinity does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we always have to deal with a finite number of trials.
>
> (2) The frequency concept applied to a large but finite number of trials does not admit a rigorous formal exposition within the framework of pure mathematics.
>
> I still maintain the first of the two theses mentioned above. As regards the second, however, I have come to realise that the concept of random distribution of a property in a large finite population can have a strict formal exposition. In fact, we can show in sufficiently large populations the distribution of the property may be such that the frequency of its occurrence will be almost the same for all sufficiently large sub-populations, when the law of choosing these is sufficiently simple. Such a conception in its full development requires the introduction of a measure of the complexity of an algorithm. I propose to discuss this question in another article. In the present article, however, I shall use the fact that there cannot be a very large number of simple algorithms.

Six years later [1969a] he wrote

> (1) The fundamental concepts of information theory can, and must, be substantiated without recourse to probability theory, and in such a way that the concepts of *entropy* and *quantity of information* are applicable to individual objects;
>
> (2) the concepts of information theory thus introduced may be the basis for a new conception of the notion *random* corresponding to the natural assumption that randomness is the absence of regularity.

To these it is proper to add another of ANK's remarks [1983d]:

> The applications of probability theory can be put on a uniform basis. It is always a matter of consequences of hypotheses about the impossibility of reducing in one way or another the complexity of the description of the objects in question. Naturally, this approach to the matter does not prevent the development of probability theory as a branch of mathematics being a special case of the general measure theory.

Let us now try to catch the gist of ANK's new approach, viewed here for the sake of simplicity in a typical 'context', that of Lebesgue measure on the Borel subsets of $\{0, 1\}^\infty$. (In the language of the *Grundbegriffe* we should have said that we have in mind an infinite sequence of Bernoulli trials with individual chance $= \frac{1}{2}$.) In fact it is characteristic of the new theory, just as it was of the old, that we have to indicate *what we are trying to model* by referring to a triple $(\Omega, \mathscr{F}, \mu)$, with $\mu(\Omega) = 1$, in the usual way. In what follows, reference to the model will be indicated by a reference to 'the context'.

The first step is to introduce four special sets of infinite (0, 1)-sequences ω to be called T, C, KS, and CS. For the precise definitions of these see [1987c]. The 'definitions' given here will be informal only, and they omit essential details that are far beyond our present scope.

A given infinite (0, 1)-sequence $\omega = (\omega_0, \omega_1, \ldots)$ will be a member of T (and is then called a *typical* sequence) if and only if it belongs to every subset of $\{0, 1\}^\infty$ that *effectively* has measure 1. It is a theorem of P. Martin-Löf that T, defined in this way, has itself effectively measure 1 (that is, that it is the least such subset). Here 'effective' refers to the explicit algorithmic basis of the whole approach. I will not even attempt to indicate the basis of that here.

A given infinite (0, 1)-sequence $\omega = (\omega_0, \omega_1, \ldots)$ will be a member of C (and is then called a *chaotic* sequence) if and only if its initial n-segments $\omega^n = (\omega_0, \omega_1, \ldots, \omega_{n-1})$ have a 'complexity' (or 'entropy') $K(\omega^n)$ that grows, as n increases, *at the fastest possible rate*. This definition presumes that we have given a prior definition of the 'optimal monotone complexity' $K(\theta)$ (sometimes called $KM(\theta)$) of a *fixed finite* object θ, again relative to the 'context'. The basic idea is that this complexity is essentially just the length of the shortest possible description of θ. Note that the word 'shortest' refers to the 'context', so that in our present example C, like T, is to be viewed from the standpoint of one interested in the classical concept of Bernoulli trials with individual probability p (here $= \frac{1}{2}$), and in minimising the length of the description we are allowed to be influenced by this. I shall say no more about the definition of 'optimal monotone complexity' than to remark that we can always suppose the length of the description to be less than or equal to n, and that 'growing at the fastest possible rate' is to mean that $K(\omega^n) \geqslant n - c$ for some positive c independent of n but perhaps depending on ω. Different 'optimal monotone complexities' and the associated minimal descriptions of a finite object θ relative to the 'context' will in general lead to different values of $K(\theta)$, but a fundamental theorem asserts that any two such optimal monotone complexities K^* and K^{**} always satisfy an inequality of the form

$$|K^*(\theta) - K^{**}(\theta)| \leqslant c^{*,**},$$

where $c^{*,**}$ does not depend on θ, and this ensures that there is no ambiguity in the definition. (As explained before, it is the references to 'context' that unambiguously lock the discussion onto what the classical probabilist would recognise as a specific model.)

We now have a theorem (due to L. A. Levin and C. P. Schnorr) saying that for a given 'context' the sets T and C are *the same*. That is, a *given* infinite (0, 1)-sequence ω is either (1) typical and chaotic, or (ii) non-typical and non-chaotic.

Finally the definitions of the sets CS and KS of infinite (0, 1)-sequences serve to remind us of what the 'contextual' probability model was. It is formulated in language similar to that used by von Mises when describing his 'collectives'. If we used the Church reformulation of that, we would have to say that $\#(\omega^n)/n \to p$ as $n \to \infty$ (where $\#(\omega^n)$ is the sum of the components of ω^n, and where in our present example $p = \frac{1}{2}$), *and that this convergence holds in every effectively selected subsequence.* That, however, would tell us that the given infinite (0, 1)-sequence ω belongs to set CS (C for Church). We want the necessary and sufficient condition for ω to belong to a different set, KS (K for Kolmogorov), which is to be a subset of CS, and it is defined by requiring the same convergence condition to hold even when in forming the selected subsequences we are allowed at each stage to select *any symbol in the sequence that has not already been chosen* – that is, we are allowed (*effectively*) to 'dodge about' when selecting new terms.

We then have a second theorem saying that

$$T = C \subset KS \subset CS.$$

Here the inclusion $C \subset CS$ is strict (M. van Lambalgen and D. Loveland).

Accordingly we can use C to provide an environment in which to do classical probability with a new – an *entropic* – motivation. It is natural to ask if the *first* inclusion in the above displayed formula could be shown to be an equality. In one of his publications [1969a] ANK announced results suggesting that KS is strictly larger than C, but the proof of these assertions has since been lost, and so that question is still open. (Dr Razborov now tells me that A. Shen has proved this result.)

Another open question asks whether an element of KS stays in KS if we apply a second Kolmogorov selection process to it.

To practical probabilists many other questions will spring to mind, and it is too early to expect conclusive answers to all of them. The time is, I suppose, not yet ripe for an entropy-theoretic reworking of the *Grundbegriffe*, but we may perhaps hope to see this done in the near future.

Already the fact that $C \subset KS$ is enough to make it plain that in principle we should be able to rebuild probability theory starting with the infinite (0, 1)-sequences in C as a basis, so using entropic rather than probabilistic methods. Before he died, ANK was convinced that this must be so, and indeed he knew that to some extent it had been carried out. It was therefore especially fitting that the Kolmogorov–Uspenskii paper [1987c] that we have been following, which was delivered (by Uspenskii) as the opening lecture of the First World Congress of the Bernoulli Society in Tashkent, should be immediately followed in its subsequent publication in *Teoriya Veroyatnostei* by a remarkable paper by V. G. Vovk that exemplifies in a triumphant manner the success of this part of the Kolmogorov programme.

This is nothing less than *an entropic proof of the classical law of the iterated logarithm for a FIXED chaotic infinite* (0, 1)-*sequence*. As the LIL theorem first emerged in the context of number theory (see for example Feller [13]), this is natural enough, but one wonders what G. H. Hardy would have thought of it.

Let us use the notation

$$\mathrm{LIL}^*(\omega) = \limsup \frac{\#(\omega^n) - \frac{1}{2}n}{\sqrt{(n \log\log n)}},$$

so that the classical theorem asserts that $\mathrm{LIL}^*(\omega) = 1/\sqrt{2}$,

Let us now introduce the following terminology: we shall say that $\omega \in \{0, 1\}^\infty$ is *chaotic up to a discrepancy* $f(n)$ when

$$n - K(\omega^n) \leqslant f(n) + O(1),$$

and that it is *chaotic* if the condition holds with $f(n) = O(1)$. Then Vovk's version of the LIL theorem is that

$$\mathrm{LIL}^*(\omega) = \frac{1}{\sqrt{2}} \text{ if } \omega \text{ is chaotic up to a discrepancy } o(\text{loglog } n).$$

Accordingly we get an entropic version of the classical LIL-theorem with an unexpectedly light assumption, because the Vovk assumption for LIL is much weaker than mere chaoticity (= membership of C). Vovk's theorem indeed generalises the classical LIL theorem in two ways. In the first place, it holds for any *fixed* infinite binary sequence provided that a suitable degree of chaos prevails, and in the second place, the critical chaotic condition is substantially weaker than the one that might have been expected. But Vovk's techniques in fact yield much more than this. He is also able to describe what it is that replaces the classical LIL behaviour when the given (0, 1)-sequence satisfies any one of a variety of yet weaker near-chaos conditions.

Vovk then turns to two other classical limit theorems and shows, for example, that the strong law of large numbers holds for an individual infinite (0, 1)-sequence ω if it is chaotic up to a discrepancy $o(n)$, and that it fails for some of the ω that are chaotic up to a discrepancy εn. Also he shows that 0 and 1 each recur infinitely often if the given ω is chaotic up to a discrepancy $(\frac{1}{2}-\varepsilon)\log_2 n$, but that there exists an 'ultimately constant' sequence ω that is chaotic up to a discrepancy $(2+\varepsilon)\log_2 n$.

So we can now assert classical probability limit theorems for suitably *nearly* chaotic *individual* infinite (0, 1)-sequences, and also we can now classify such theorems by the degree of chaoticity required. This last aspect of Vovk's work reminds one of the concept of 'depth' in number theory, not to mention the 'Infinitärkalkül' of Du Bois Reymond. Hardy would indeed have been interested! I hope that work has already begun on computing the discrepancy $f(n)$ for some of the more interesting infinite (0, 1)-sequences that occur in the classical theory of numbers.

Another very striking recent result is that of E. A. Asarin (whose paper immediately follows that of Vovk). He has shown that the analogue of the 'T = C'-theorem holds in the very different 'context' of brownian paths.

In concluding this mini-review of ANK's later work I must stress again that the results described were obtained by a large and internationally diverse school, but it was a school in effect dedicated to making explicit the perceptions of Kolmogorov.

Mention must be made here of the use of quantitative complexity in inferential statistics. To take only one example of this, when two competing explanatory theories have been proposed with reference to a given set of observed data, it is usually thought desirable to penalise the more complicated explanation in some quantitative way, when carrying out a statistical test to decide which explanation to adopt. Some numerical measure of complexity will be required for that purpose, and the choice and use of this is related to ANK's programme just because it involves the definition of complexity for a given finite object. A stimulating discussion of a wide range of such questions will be found in the record of a Symposium organised by the Royal Statistical Society (*Journal*, Series B, 1987) and built around papers by P. R. Freeman, J. Rissanen, and C. S. Wallace presenting complexity-based approaches to a variety of statistical problems.

At all stages of his career ANK seems to have been busy simultaneously on a multiplicity of fronts, and this was especially so towards the end of his life. Thus during the decade of 'complexity' he was also actively developing new limit theorems of the classical type with his younger colleagues, and occupied with his growing interest in mathematical education, taking very heavy responsibilities in connexion with one of the special schools for gifted children sponsored by the Moscow State University. To this school he devoted a major proportion of his time over many years, planning syllabuses, writing textbooks, spending a large number of teaching hours with the children themselves, introducing them to literature and music, joining in their recreations and taking them on hikes, excursions, and expeditions. There are those who shudder when such schools are mentioned, and protest about excessive pressure, one-sided development, and so forth. I am persuaded that these criticisms are groundless, and founded on ignorance. ANK sought to ensure for these children a broad and natural development of the personality, and it did not worry him if the children in his school did not become mathematicians. Whatever profession they ultimately followed, he would be content if their outlook remained broad and their curiosity unstifled. Indeed it must have been wonderful to belong to this extended family of Andrei Nikolaevich. See him talking to some members of it (Plate II).

A wise man has remarked that every mathematician has his own personal view of Kolmogorov. I have attempted to portray my own, but the numerous obituaries now being written will enable readers to sample other perspectives, and to build up a portrait for themselves. I should add that the hard facts, the bibliographic details, and the three photographs included in this impressionistic essay were given to me by my friend Albert Shiryaev, to whom I am most grateful. His own much longer study will be published in the journal *Annals of Probability* and will undoubtedly prove to be the authoritative record of the life and work of this extraordinary man.

Those who want a glimpse of Kolmogorov's personality will find it of interest to read the two sets of reminiscences published by Aleksandrov [**1**] and by Kolmogorov himself [1986b]. These contain fascinating records of expeditions to the mountains (see Plate III), and record the history of their long friendship.

It could go without saying that Kolmogorov received numerous honours. He was a Hero of Socialist Labour, and was awarded seven Orders of Lenin in addition to many other distinctions. But perhaps the greatest honours conferred upon him in his own country were the Lenin Prize (1965) and the Lobachevskii Prize (1987). At the time of his death he was clearly recognised there, as also here, to be one of the greatest mathematicians of all time.

From abroad came 26 honorary doctorates and honorary memberships of learned societies, as well as a Balzan Prize and a Wolf Prize. In particular he became an Honorary Member of the London Mathematical Society in 1959, an Honorary Fellow of the Royal Statistical Society in the same year, and a Foreign Member of the Royal Society in 1964.

He directed the studies of nearly 70 research pupils, of whom some, as for example

Martin-Löf and Rényi, came from other countries, but mostly his pupils were from the USSR, many of these later becoming Members or Corresponding Members of the Soviet Academy of Sciences.

In a moving last message to his research pupils, quoted in the memoir by Shiryaev, he laid upon them the responsibility of continuing his work for the better education of young children. It seems certain that here too we all have much to learn from his example.

We can only guess how Kolmogorov will be regarded by future generations. Which was the most significant: his massive combinatorial power, or his penetrating insight? Or should these be regarded as two aspects of a single gift?

References

1. P. S. ALEKSANDROV, 'Pages from an autobiography', *Uspekhi Mat. Nauk* 34 (1979) 219–249, 35 (1980) 241–278.
2. M. S. BARTLETT, 'Periodogram analysis and continuous spectra', *Biometrika* 37 (1950) 1–16.
3. D. BLACKWELL, 'Another countable Markov process with only instantaneous states', *Ann. Math. Statist.* 29 (1958) 313–316.
4. N. F. BRITTON, *Reaction-diffusion equations and their applications in biology* (Academic Press, New York, 1986).
5. R. BROWN, 'A brief account of microscopical observations made... 1827 &c', *Philos. Mag.* (2) 4 (1828) 161–173.
6. P. J. DANIELL, 'Integrals in an infinite number of dimensions', *Ann. of Math.* (2) 20 (1918–19) 281–288.
7. P. J. DANIELL, 'Functions of limited variation in an infinite number of dimensions', *Ann. of Math.* (2) 21 (1919–20) 30–38.
8. P. J. DANIELL, 'Contribution to discussion', *J. Royal Statist. Soc. Suppl.* 8 (1946) 88–90.
9. DI SAN-MIN, 'Uniqueness of Doob process determined by '0+'-system', *Essays on Markov functions*, 4th Chinese Probability Colloq. (Institute of Railways, Changsha, 1984), pp. 125–136.
10. J. L. DOOB, *Stochastic processes* (Wiley, New York, 1953).
11. A. EINSTEIN, 'Uber die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen', *Ann. d. Physik* (4) 17 (1905) 549–560.
12. N. I. ERMOLAEVA, *Yarovizatsiya* 2 (1939) 79.
13. W. FELLER, *An introduction to probability theory and its applications*, Vol. I (3rd ed., Wiley, New York, 1968).
14. R. A. FISHER, *The genetical theory of natural selection* (Oxford University Press, 1930).
15. R. A. FISHER, 'The wave of advance of advantageous genes', *Ann. Eugenics* 7 (1937) 355–369.
16. M. FRÉCHET, *Recherches théoriques modernes sur le calcul des probabilités*, Vol. I (1936), Vol. II (1938) (Gauthier Villars, Paris).
17. M. L. GASPAROV et al., *Problemi teorii stiha* (Nauka, Leningrad, 1984).
18. B. HOSTINSKY, *Méthodes générales du calcul des probabilités*, Mem. Sci. Math. 52 (1931).
19. HOU ZHEN-TING, "0+'-system determines birth and death process uniquely', *Essays on Markov functions*, 4th Chinese Probability Colloq. (Institute of Railways, Changsha, 1984), pp. 103–112.
20. C. T. IONESCU TULCEA, 'Mesures dans les espaces produits', *Atti Acad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Nat.* (8) 7 (1949) 208–211 (1950).
21. D. G. KENDALL, 'Mathematical models of the spread of infection', *Math. and Computer Sci. in Biol. and Medicine*, (Med. Research Council) (HMSO, London, 1965), pp. 213–224.
22. D. G KENDALL, 'Some recent advances in the theory of denumerable Markov processes', *Trans. 4th Prague Conf. on Info. Theory etc.* (1967), pp. 11–17.
23. D. G. KENDALL and G. E. H. REUTER, 'Some pathological Markov processes etc., *Proc. Internat. Congress Math. Amsterdam* (1954) III, 377–415.
24. E. KOLMAN, 'Is it possible to prove or disprove Mendelism by mathematical and statistical methods?' *C.R. Acad. Sci. URSS* 28 (1940) 834–838.
25. P. LEVY, 'Systèmes markoviens et stationnaires. Cas dénombrable', *Ann. Sci. École Norm. Sup.* (3) 68 (1951) 327–381.
26. T. D. LYSENKO, 'In response to an article by A. N. Kolmogoroff', *C.R. Acad. Sci. URSS* 28 (1940) 832–833.
27. R. VON MISES, *Wahrscheinlichkeitsrechnung* (Fr. Deuticke, Leipzig and Vienna, 1931).
28. O. NIKODYM, 'Sur une généralisation des intégrales de M. Radon', *Fundamenta. Math.* 15 (1930) 131–179.

29. D. ORNSTEIN, 'The differentiability of transition functions', *Bull. Amer. Math. Soc.* 66 (1960) 36–39.
30. G. E. H. REUTER, 'On Kendall's conjecture concerning '0+'-equivalence of Markov transition functions', *J. London Math. Soc.* (2) 35 (1987) 377–384,
31. N. WIENER, *Extrapolation, interpolation, and smoothing of stationary time series* (Wiley, New York, 1949).
32. A. A. YUSKEVICH, 'On differentiability of transition probabilities of homogeneous Markov processes with a countable number of states', *Ucenye Zapiski MGY* 186 (Mat.) 9 (1959) 141–159.

KOLMOGOROV'S WORK ON TURBULENCE

G. K. BATCHELOR

The name Kolmogorov has different associations for different mathematical and scientific communities. For those interested in the turbulent motion of fluid, Kolmogorov – whom they think of as *their* Kolmogorov – will always be remembered for the theory of 'local isotropy', or universal equilibrium, of the small-scale components of fluid motion that he put forward in 1941. This powerful theory, which has been found extremely useful in a wide variety of physical contexts, was published (in English) in two short notes [1941d, g] in *Comptes Rendus de l'Académie des Sciences de l'URSS* which, remarkably, found their way into western university libraries during World War II. These papers are unusual for a mathematician in that they contain no mathematics to speak of, dimensional analysis and elementary probability concepts being all that is required. They are essentially a statement of two hypotheses, justified heuristically, and the style is that of an axiomatic theoretical physicist. Kolmogorov published four other similar short notes on turbulence [1941f, 1942b, 1949h, 1962f], the last being a reconsideration and refinement of the universal equilibrium theory. All these papers are unusual, original and penetrating, but the two in 1941 on the universal equilibrium theory have had by far the greatest impact on the study of turbulence and will be described first.

The basis of Kolmogorov's two hypotheses was the notion of kinetic energy 'cascading' from components of the fluctuating turbulent motion with large length scales to components with smaller length scales as a consequence of nonlinear inertial interaction of different components, a notion which was familiar from previous work by G. I. Taylor and L. F. Richardson in particular. Kolmogorov recognised that, as the Reynolds number of the turbulent motion is increased, the smallest, viscosity-dominated, length scale present in the flow decreases, thereby increasing the number of steps in the cascade; and he regarded as plausible the assumptions, first that the cascade has an accelerating character in that the transfer of energy from components with length-scale l, say, to those with length-scale $\frac{1}{2}l$ occurs in a time which diminishes with l, and second that the transfer from one scale to another is accompanied by some statistical decoupling of the components involved. This physical picture of the turbulence suggested to Kolmogorov the premise that the components of motion with length scales small compared with the scale L characteristic of the components of motion containing most of the kinetic energy are effectively independent of the components on length-scales near L, and so are asymptotically statistically homogeneous in space and time and isotropic, regardless of the way in which the turbulence is being generated and regardless of its large-scale statistical properties.

This extremely powerful premise represents the essence of the Kolmogorov theory. It was latent in earlier discussions of observational data, but had not been formulated nor seen so clearly. From a practical point of view it has the weakness of referring only to components on small length scales, which normally make negligible contributions to the rates of transfer of momentum and mass in inhomogeneous flow fields, but about these small-scale components it says a great deal.

If the above premise is accepted, there arises the question, what determines the properties of the components of motion with small length scales involved in this universal equilibrium? This question was answered by Kolmogorov's two 'similarity hypotheses'. The first of these hypotheses states that the statistical properties of the small-scale components of the motion are uniquely determined by just two parameters, one being the kinematic viscosity of the fluid (ν) which is relevant to the dissipation of energy at the very smallest scales, and the other the mean rate at which energy is transferred to the universal-equilibrium range of length scales from larger scales per unit mass of fluid (ε). The idea here is that the large-scale properties of the turbulence are relevant only insofar as they are the source of kinetic energy which is transferred, at the rate ε, to the smaller-scale components. Moreover, since the small-scale components of the motion are statistically steady, the mean rate at which energy is put in at one end of the universal-equilibrium range is equal to the mean rate at which it is taken out at the other end by viscous dissipation, showing that

$$\varepsilon = \frac{1}{2}\nu \sum_{i,j} \left\langle \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)^2 \right\rangle$$

where u_i is the local velocity of the fluid.

This first hypothesis imposes restrictions on the functional forms of statistical properties of the small-scale components of the motion. For instance, for the second moment of the difference between the fluid velocities at two points separated by the vector $\mathbf{r}$ we have on dimensional grounds

$$\langle \{u_d(\mathbf{x}+\mathbf{r}) - u_d(\mathbf{x})\}^2 \rangle = (\nu\varepsilon)^{\frac{1}{2}} F(r/\eta)$$

for $r \ll L$, where d denotes a component parallel to $\mathbf{r}$, r is the scalar magnitude of $\mathbf{r}$, and $\eta = (\nu^3/\varepsilon)^{\frac{1}{4}}$. The length η is evidently a measure of the length scale at which viscous dissipation occurs, and the function F is a universal form. It may also be shown from the statistical isotropy of the small-scale motion that the second moment of any other component of the velocity difference is determined by this same function F.

The second similarity hypothesis states that, if the Reynolds number of the turbulence is so large that values of r such that $\eta \ll r \ll L$ exist, the properties of the components of motion associated with length scales in this range depend only on ε. This hypothesis strengthens the restrictions, and again it follows on dimensional grounds alone that F must be proportional to $(r/\eta)^{\frac{2}{3}}$, corresponding to

$$\langle \{u_d(\mathbf{x}+\mathbf{r}) - u_d(\mathbf{x})\}^2 \rangle = C(\varepsilon r)^{\frac{2}{3}}$$

for $\eta \ll r \ll L$, where C is a universal constant. Similar deductions may be made for other moments of velocity differences; and any dimensionless ratio of powers of these moments is necessarily an absolute constant when r lies in this range in which only inertia forces act.

This expression for the second moment of the velocity difference was the first definite quantitative prediction in the study of turbulence. Kolmogorov compared the predicted form of the second moment with the very limited wind-tunnel data available in 1941, but the measurements were too sparse to provide any real test of the variation as $r^{\frac{2}{3}}$ over the rather small range of values of r to which it applies. Techniques for the measurement of fluctuating velocities have improved greatly since then, and it has also been realised that it is preferable to compare measurements with predictions of the energy spectrum function, which is the Fourier transform of the second moment of the velocity, because the wave-number range corresponding to $\eta \ll r \ll L$ enlarges indefinitely as the Reynolds number is increased. The spectral density should vary as $\varepsilon^{\frac{2}{3}}\kappa^{-\frac{5}{3}}$ (where κ is the magnitude of the vector wave number) provided $L^{-1} \ll \kappa \ll \eta^{-1}$, according to the Kolmogorov theory, and this has now been supported by many independent sets of measurements, over several wave-number decades in the case of measurements at very large Reynolds number.

By an extraordinary coincidence the essential ideas of the universal equilibrium theory were put forward independently a short time later by two other people, L. Onsager in 1945 in USA, and C. F. von Weizäcker in Germany in 1945 although not published until 1948. They both showed that the spectral density should vary as $\kappa^{-\frac{5}{3}}$ for a restricted range of wave-number magnitude. However, the clearest formulation of the ideas was undoubtedly that of Kolmogorov, and it was also both more precise and more general.

A few years later Kolmogorov [1949h] made what proved to be a typical application of the universal equilibrium theory to a physical problem. In these applications one first identifies a physical process in which the components of motion with length scales in the equilibrium range play a significant part, and then the effect of these small-scale components is represented analytically by the appropriate dimensional combination of ε and ν. Kolmogorov's problem concerned the tendency for turbulent motion of a dispersion of drops of one liquid in a second liquid to deform and perhaps to break up the drops. The largest size of drop for which surface tension at the interface can hold the drop together against the deforming tendency of the turbulence will be that for which the surface-tension stress T/a (where a is the drop radius) is comparable in magnitude with the variation of stress due to the turbulence over the outer surface of the drop. When $\eta \ll a \ll L$, this variation of stress over the drop surface is primarily inertial and is determined mainly by the components of the motion with length scales near a, and so on dimensional grounds is proportional to $\rho(\varepsilon a)^{\frac{2}{3}}$. Equating the two stresses then gives the useful prediction that a is of order $(T/\rho)^{\frac{3}{5}}\varepsilon^{-\frac{2}{5}}$. Many other similar applications of the universal equilibrium theory have been made.

But the 1941 universal equilibrium theory was not perfect. At an international conference on the mechanics of turbulence at Marseille in 1961 Kolmogorov announced that 'quite soon after' these ideas originated 'Landau noticed that they did not take into account a circumstance which arises directly from the assumption of the essentially accidental and random character of the mechanism of transfer of energy from the coarser vortices to the finer'. Landau's point was that the local and instantaneous rate of energy dissipation per unit mass may be expected to have an increasingly spotty distribution as the Reynolds number of the turbulence is increased and that its variance will increase without limit. Other properties of the local small-scale components of the motion will likewise exhibit large fluctuations, and any relation between statistical quantities which are affected by fluctuations about the mean will consequently be dependent on the Reynolds number and so cannot be truly universal. For example, the dimensionless ratio of the nth moment ($n > 2$) of the local velocity gradient to the $\frac{1}{2}n$-power of the second moment is a parameter of the small-scale components and so should be a universal constant according to Kolmogorov's first similarity hypothesis, but may be expected to depend on the Reynolds number in the light of Landau's comment. Measurements of such ratios of moments have been made, and they confirm that there is indeed a dependence on Reynolds number, the dependence being an increase in magnitude with Reynolds number which becomes more rapid as n is increased.

Further progress clearly required information about the probability distribution of the local rate of energy dissipation. At this same conference Kolmogorov [1962f] went on to report that his colleague and former student A. M. Obukhov had proposed, as 'a simplified scheme', that the logarithm of the average of the instantaneous rate of dissipation over a sphere of radius r in the fluid ($\ln \varepsilon_r$) has a normal distribution when L/r is large and that the variance of $\ln \varepsilon_r$ is given by

$$A+\mu \ln (L/r),$$

where A depends on the large-scale features of the motion and μ is a universal constant. Kolmogorov thereupon modified his two similarity hypotheses to allow for dependence on the Reynolds number, and adopted Obukhov's specific suggestion as a third hypothesis. The explicit expression for the energy spectral density now becomes

$$E(\kappa) = C\varepsilon^{\frac{2}{3}}\kappa^{-\frac{5}{3}}(\kappa L)^{-\mu/9} \quad \text{for} \quad L^{-1} \ll \kappa \ll \eta^{-1}.$$

It appears from a number of experiments made in recent years that this expression fits the data a little better with μ having a value between 0.2 and 0.5 than with $\mu = 0$, although the difference is slight. The effect of dissipation fluctuations on the moments of the velocity gradient is stronger, and here too there is reasonable agreement between observations and the consequences of the third hypothesis. The properties of the small-scale components of turbulent motion, and in particular the intriguing 'intermittency' in the spatial and temporal distributions of vorticity and dissipation, are still the subject of discussion and research, nearly 50 years after Kolmogorov announced his universal equilibrium theory.

Finally, there is a paper [1942b] about quite different questions which is less profound but remarkably prescient. Here Kolmogorov had the very practical purpose of establishing approximate equations which would allow calculation of some of the important parameters of flow fields in which turbulence is generated by a mean shearing motion of the fluid and is dependent on position in the fluid, as for instance in steady mean flow along a tube of circular cross-section. It is not possible to obtain a closed exact set of governing equations for mean quantities from the Navier–Stokes equation of motion, because the number of velocity moments needed to specify the fluctuating motion statistically is not finite and as a consequence of the nonlinearity of this equation the moments of different order are interdependent. Kolmogorov proposed instead to choose a small number of physically significant quantities (such as the local mean kinetic energy of the fluctuating motion per unit mass of fluid) which would be the dependent variables in a corresponding set of approximate equations obtained by operating on the Navier–Stokes equation in various ways before averaging, an idea on which Prandtl was working simultaneously in Germany. The difficult part of the plan is to represent, by an intuitive appeal to some physical picture of the processes at work, each of the quantities arising in these equations in terms of the chosen dependent variables. It calls for inspired guessing and a judicious compromise between simplicity of the equations and accuracy of the representation. Kolmogorov chose three position-dependent variables, the mean velocity, the mean square of the velocity fluctuation, and a certain frequency related to the mean rate of strain. His proposed approximate equations need not be reproduced here, because they have not stood the test of time and it is the approach that is significant. The equations are inevitably non-linear, and Kolmogorov noted that their solution 'presents great difficulties'. Numerical solution was not feasible in 1942, but it has become so in recent years, and the approach to turbulent shear flow suggested by Kolmogorov is now the basis of a vast amount of work, known as 'turbulence modelling', directed towards the solution of practical flow problems in mechanical, aeronautical, hydraulic and chemical engineering.

KOLMOGOROV'S WORK ON PROBABILITY, PARTICULARLY LIMIT THEOREMS

N. H. Bingham

1. *Foundations and the 'Grundbegriffe'*

In Hilbert's problem list of 1900 one finds (Problem 6) 'To treat in the same manner, by means of axioms, those physical sciences in which mathematics plays an important part; in the first rank are the theory of probability and mechanics'.

The basis for a modern and rigorous treatment of probability was laid by Lebesgue around 1901–4 in his work on measure theory and integration. In particular, this provided the language in which results on almost-sure convergence could be formulated, an early example being Borel's 'normal number' theorem of 1909.

It is perhaps surprising that some three decades were needed before the successful synthesis of these ideas took place. One necessary preliminary was the freeing of measure theory from the geometrical aspects associated with its being developed first in Euclidean space; a major influence here is the work of Fréchet. Another was the completion by Nikodým in 1930 of the 'Radon–Nikodým theorem', begun by Radon in 1913; this was an essential ingredient in a successful treatment of conditioning.

The flourishing Soviet school of analysts including Luzin, Suslin and others did much to develop the 'metric theory of functions', and Kolmogorov, a pupil of Luzin, was well placed to turn the measure-theoretic background he acquired here to probabilistic problems, in which he became interested following his first collaboration with Khinchin [1925f]. The result was his classic book on foundations of 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, which essentially inaugurated the modern era of probability theory.

Chapter I (in which Kolmogorov credits Fréchet as being the first to identify probability, expectation, random variable,... with measure, integral, measurable function,...) is introductory. In Chapter II we find σ-algebras and the key axiom of countable additivity, completing the now-standard Kolmogorov axioms. Chapter III develops distribution functions in one and several dimensions, and then (§4) in infinitely many. Here we find the key 'Daniell–Kolmogorov theorem' (originally due to P. J. Daniell in 1918), passing from an appropriately consistent infinite set of 'finite-dimensional distributions' to one 'infinite-dimensional distribution', or stochastic process. In §5 one finds convergence in probability and with probability one. Chapters IV and V cover expectation and conditioning. In Chapter VI we find the weak and strong laws of large numbers (considered further below), and in the Appendix the Kolmogorov zero-one law.

It is difficult to overstate the impact of the *Grundbegriffe* on the development of the subject; essentially the history of probability theory splits in 1933 between 'pre-

Kolmogorov' and 'Kolmogorov'. We note in passing that history might have turned out rather differently; very important measure-theoretic work had already been done by Paul Lévy – for instance, in his book [**17**] of 1925 one finds the modern machinery of characteristic functions. In his autobiography [**20**, pp. 67–68] Lévy writes poignantly of his realisation, immediately on seeing the *Grundbegriffe*, of the opportunity which he himself had neglected to take. A rather different perspective is supplied by the eloquent writings of Mark Kac ([**16**], and preface to [**15**]) on the struggles that Polish mathematicians of the calibre of Steinhaus and himself had in the 1930s, even armed with the *Grundbegriffe*, to understand the (apparently perspicuous) notion of stochastic independence. This period resulted in much important probabilistic work with a Polish flavour – for instance, the delightful book by Kac [**14**] on independence, and the work of Marcinkiewicz and Zygmund [**24**].

2. *Weak limit theorems; Gnedenko and Kolmogorov*

Lévy's book [**17**], and the continuity theorem for characteristic functions therein, provided a powerful new technique for proving weak (or distributional) limit theorems. Suppose for instance that $X, X_1, X_2, \ldots$ are independent and identically distributed with mean μ and characteristic function ϕ. From

$$\mathrm{E}\exp\left\{it\sum_{k=1}^{n} X_k/n\right\} = \phi(t/n)^n = (1+i\mu t/n+o(1/n))^n \to e^{i\mu t}$$

and the continuity theorem, one obtains the 'weak law of large numbers'

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to \mu \quad (n\to\infty) \quad \text{in probability,}$$

due to Khinchin in 1929. Kolmogorov considers the non-identically distributed case in his classic paper [1928d] and its sequel [1929a]. In particular (Satz XII) he refines Khinchin's weak law in the identically distributed case, showing that there exist constants c_n with

$$\frac{1}{n}\sum_{k=1}^{n} X_k - c_n \to 0 \quad \text{in probability}$$

((X_n) is 'stable') if and only if

$$n\,P(|X|>n)\to 0;$$

then c_n can be taken as $\int_{-n}^{n} y\,dF(y)$ (writing F for the law of X). The final result here is that of Ehrenfeucht and Fisz of 1960, giving the equivalence of

(i) the characteristic function ϕ is differentiable at the origin, and $\phi'(0) = i\mu$,
(ii) $x\mathrm{P}(|X|>x)\to 0$ and $\int_{-x}^{x} y\,dF(y)\to\mu$ $(x\to\infty)$,
(iii) $1/n\sum_1^n X_k\to\mu$ $(n\to\infty)$ in probability.

Slightly more general is the question of 'relative stability': the existence of constants a_n with

$$\sum_{k=1}^{n} X_k/a_n \to 1 \text{ in probability,}$$

a classical instance being that of the St Petersburg game.

The central limit theorem may be handled similarly. When the X_n have variance σ^2, writing ϕ_0 for the characteristic function of $X - \mathrm{E}X$ one has

$$\mathrm{E}\exp\left\{it\sum_{k=1}^{n}(X_k-\mu)/(\sigma\sqrt{n})\right\} = \phi_0\left(\frac{t}{\sigma\sqrt{n}}\right)^n = \left(1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)^n \to e^{-\frac{1}{2}t^2},$$

and (writing Φ for the standard normal or Gaussian law) concludes that

$$\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(X_k-\mu)\to\Phi \quad (n\to\infty) \text{ in distribution}$$

by the continuity theorem. This central limit problem may be generalised: consider 'triangular arrays' $(X_{nk}: 1 \leqslant k \leqslant k_n, n = 1, 2, \ldots)$, with terms in the nth row mutually independent, and individual terms negligible as $n \to \infty$. What are the possible limit laws for $\sum_{k=1}^{k_n} X_{nk}$, and what are the conditions for convergence? The possible limits are the infinitely divisible laws (those which, for each $n = 1, 2, \ldots$, are an nth convolution power), characterised by the Lévy–Khinchin formula (due in the finite-variance case to Kolmogorov [1932b, c], and in the general case to Lévy in 1934/35, Khinchin in 1937).

Kolmogorov's second profoundly influential contribution to the textbook literature of probability theory is his monograph with B. V. Gnedenko [1949a], published in Russian in 1949 and translated into English (and annotated) by K. L. Chung in 1954. Chapter 3 contains a thorough discussion of the infinitely-divisible laws, Chapter 4 of conditions for convergence to them, and Chapter 5 of conditions for convergence to normal, Poisson or degenerate laws (this last covering the weak law of large numbers, relative stability etc.). Chapters 7–9 cover the identically distributed case (stable laws and their domains of attraction, Berry–Esseen theorem, local limit theorems, etc.).

Something of the power and scope of [1949a], as well as its style, is aptly summarised by its translator, Chung, in his preface: '...a certain amount of mathematical maturity, perhaps a touch of single-minded perfectionism, is needed to penetrate the depth and appreciate the classic beauty of this definitive work'. Of its central theme, Chung remarks again, in the preface to his own book [3], that it 'has been called the "central problem" of classical probability theory. Time has marched on and the centre of the stage has shifted, but this topic remains without doubt a crowning achievement'.

One should note the debt that Gnedenko and Kolmogorov owes to its predecessor, Paul Lévy's classic [**18**] of 1937. Indeed, Lévy emphasised the dominant part that the question of conditions for convergence to normality played throughout his life as a probabilist; see his autobiography [**20**] and obituary [**22**]. As for its successors: all the modern classics treat at least part of this material; we single out the monographs of Ibragimov and Linnik [**11**] and Petrov [**25**] as perhaps closest in scope. Though much of the book's content is now available in more streamlined or easily digestible form elsewhere, it remains a powerful source of inspiration and technique.

3. *Strong limit theorems*

Kolmogorov's first work on strong limit theorems – indeed, on probability theory – is his paper with Khinchin of 1925 on convergence of random series [1925f]. He returns to the subject in [1928d], Satz VII, obtaining the criterion for convergence ('three-series theorem'). In particular, for independent X_n, convergence of $\sum \operatorname{var} X_n$

and $\sum \mathrm{E}X_n$ imply almost-sure convergence of $\sum X_n$. The subject was pursued further by Lévy in his equivalence theorem **[18]**: for X_n independent, convergence of $\sum X_n$ in distribution, in probability and with probability one are equivalent.

In [1928d] one finds the important tools known as the Kolmogorov inequalities. The first (upper, or maximal) inequality

$$\mathrm{P}\left(\max_{k\leqslant n}\left|\sum_{j=1}^{k}(X_j-\mathrm{E}X_j)\right|>\varepsilon\right)\leqslant\varepsilon^{-2}\sum_{1}^{n}\operatorname{var}X_k$$

gives a powerful generalisation of Chebychev's inequality (the case $n = 1$); the second (lower) inequality is more complicated.

In [1930a] Kolmogorov obtains the following strong law of large numbers: if the X_n are independent with mean 0 and variance σ_n^2, and

$$\sum \sigma_n^2/n^2 < \infty,$$

then

$$\frac{1}{n}\sum_{k=1}^{n} X_k \to 0\ (n\to\infty) \text{ almost surely and in } L^2.$$

Conversely, if $\sum \sigma_n^2/n^2 = \infty$ the above convergence fails for some (X_n).

Kolmogorov returns to this subject in Chapter VI of the *Grundbegriffe*, where he states without proof the definitive result, Kolmogorov's strong law of large numbers: if $X, X_1, X_2, \ldots$ are independent and identically distributed,

(i) $$\mathrm{E}|X|<\infty \ \&\ \mathrm{E}X=\mu \Rightarrow \frac{1}{n}\sum_{k=1}^{n}X_k\to\mu \text{ a.s.},$$

(ii) $$\mathrm{E}|X|=\infty\Rightarrow\frac{1}{n}\left|\sum_{k=1}^{n}X_k\right|\to\infty \text{ a.s.}$$

One may combine these:

$$\mathrm{E}|X|<\infty \ \&\ \mathrm{E}X=\mu \Leftrightarrow \frac{1}{n}\sum_{k=1}^{n}X_k\to\mu \text{ a.s.}$$

Full proofs are given (in the greater generality of L^p, $0 < p < 2$) by Marcinkiewicz and Zygmund in 1937 **[24]**.

Kolmogorov's strong law is a supremely important result, as it captures in precise form the intuitive idea (the 'law of averages' of the man in the street) identifying probability with limiting frequency. One may regard it as the culmination of 220 years of mathematical effort, beginning with J. Bernoulli's *Ars Conjectandi* of 1713, where the first law of large numbers (weak law for Bernoulli trials) is obtained. Equally, it demonstrates convincingly that the Kolmogorov axiomatics of the *Grundbegriffe* have captured the essence of probability.

Kolmogorov's second major contribution to strong limit theorems is his law of the iterated logarithm (LIL) of 1929 [1929d]. The LIL was first obtained by Khinchin in 1924 for Bernoulli trials. Kolmogorov extended Khinchin's result to general distributions, not necessarily identical. For independent X_n, write S_n for $\sum_1^n X_k$, B_n for the variance of S_n, and assume $B_n\to\infty$. Kolmogorov showed that under the almost-sure bound

$$X_n = o((B_n/\log\log B_n)^{\frac{1}{2}}) \text{ a.s.}$$

one has

$$\limsup\ [\liminf]\ S_n/(2B_n\log\log B_n)^{\frac{1}{2}} = 1\ [-1] \text{ a.s.}$$

Both the result, and the method of proof (Kolmogorov's exponential bounds) have had a great influence on later work. We note that the almost-sure bound above is sharp, as was shown by Marcinkiewicz and Zygmund in 1937 [**24**].

Kolmogorov's LIL has led, in particular, to the theory of upper and lower functions and to integral tests of Kolmogorov–Erdős–Petrovskii type. For a fuller account, see, for example [**2**].

4. *From the Kolmogorov–Smirnov test to the invariance principle*

Suppose that observations $X_1, X_2, \ldots$ are made independently from a common distribution F. How can one recover the population (or parent) distribution F from a sample $(X_1, \ldots, X_n)$ of size n, in the limit as $n \to \infty$? The answer ('fundamental theorem of statistics', due to Glivenko and Cantelli, both in 1933) is that, if

$$F_n(x) := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$$

denotes the empirical distribution (random measure with mass $1/n$ on each sample point), then one has almost-sure uniform convergence of F_n to F:

$$\sup_{x \in \mathbf{R}} |F_n(x) - F(x)| \to 0 \ (n \to \infty) \text{ a.s.}$$

In this sense, the sample determines the population distribution in the limit.

To exploit this, one needs to know the rate of convergence. This was found by Kolmogorov in [1933e]: form the statistic

$$D_n := \sqrt{n} \sup |F_n(x) - F(x)|$$

(this is easy to calculate from the data, as the supremum is attained at one of the points X_k). Then

(i) D_n is 'distribution-free': its distribution is the same for all continuous F,

(ii) D_n converges in distribution as $n \to \infty$; its limit law is

$$\sum_{k=-\infty}^{\infty} (-)^k \exp(-2k^2 x^2).$$

Kolmogorov's proof was simplified by Smirnov in 1944; D_n is thus known as the Kolmogorov–Smirnov (or KS) statistic. As the 'KS limit law' above is tabulated, one thus has a non-parametric test of the statistical hypothesis that an unknown population distribution is given by a specified F. The associated theory has been very extensively developed; see the recent monograph of Shorack and Wellner [**26**].

A heuristic approach to the Kolmogorov–Smirnov theorem, identifying the limit law above with that of the supremum of a Brownian bridge, was suggested by Doob in 1949. The proof of his result was completed by Donsker in 1952, following earlier work by Erdős and Kac in 1946. In brief, Donsker's result is a functional form of the central limit theorem. Instead of a sequence of partial sums S_n scaled to converge to a normal law, one forms by piecewise linear interpolation a sequence of random continuous functions on $[0, 1]$, ξ_n say. These are shown to converge to Wiener measure (the law of Brownian motion on $[0, 1]$) in the sense of weak convergence of measures. The term 'invariance principle' is used, since the limit law does not depend on the law F of the X_n; one may thus calculate it for some simple choice of F (the coin-tossing case, say).

The Erdős–Kac–Donsker invariance principle was generalised by Kolmogorov's pupil Yu. V. Prokhorov in 1956, in the first volume of the Soviet journal *Theory of Probability and its Applications* (*TPA*). Here one finds Prokhorov's theorem identifying tightness with relative compactness. The functions space $C[0, 1]$ of continuous functions is not always the appropriate one; often one needs the space $D[0, 1]$ of functions without discontinuities of the second kind (say, right-continuous with left limits). Appropriate topologies for the study of weak convergence in D were studied by Skorokhod in 1956, again in the first volume of TPA; these were metrised by Kolmogorov [1956a]. The theory sketched above is developed at length in the influential monograph of Billingsley [**1**]. For a full account of the theory in a martingale setting, see the recent book by Jacod and Shiryaev [**12**], whose second author, A. N. Shiryaev, is again a pupil of Kolmogorov's.

5. *Stochastic processes: realisability and metric entropy*

One of the most important landmarks in the development of probability since the *Grundbegriffe* was the publication twenty years later of Doob's classic book [**4**] on stochastic processes. The Kolmogorov axiomatics were by then quite standard: we quote from Doob's preface 'Probability is simply a branch of measure theory, with its own special emphasis and field of application, and no attempt has been made to sugar-coat that fact'.

One of the distinctive problems of stochastic process theory is that of constructing (or 'realising') a stochastic process on its natural carrying space. To take the most important example, the Brownian motion or Wiener process (with time-parameter $t \in [0, 1]$, say), the Daniell–Kolmogorov theorem constructs the process on $\mathbb{R}^{[0,1]}$, but one can do much better: after discarding the complement of a Wiener-thick set one can construct the process on $C[0, 1]$ ('Brownian paths are continuous': see, for example, [**4**, VIII.2]).

Kolmogorov was one of the first to ask for criteria for a stochastic process X to be realisable on $C[0, 1]$. The classical Kolmogorov criterion (sufficient condition) for this is

$$\mathrm{E}(|X_{t+h} - X_t|^r) \leqslant c|h|^{1+s} \ (r, s > 0)$$

for some c and all t, $t+h \in [0, 1]$ (Slutskii [**27**]). This can be extended (Loève [**21**], §35]): if

$$\mathrm{P}(|X_{t+h} - X_t| \geqslant g(h)) \leqslant q(h) \to 0 \ (h \to 0) \quad \forall t,$$

then subject to suitable conditions on $g(.)$ and $q(.)$ one may realise X with (almost-surely) continuous paths, and indeed with a.s. modulus of continuity $g(.)$:

$$|X_{t+h} - X_t| \leqslant cg(h) \ \forall t, \text{ a.s.}$$

In the Brownian case, one has Lévy's modulus of continuity [**19**]:

$$\lim_{h \downarrow 0} (2h \log(1/h))^{-\frac{1}{2}} \sup_{\substack{s \in [0, 1-h] \\ t \in [0, h]}} |X_{s+t} - X_s| = 1 \text{ a.s.}$$

Suppose for simplicity that X is zero-mean Gaussian. Its structure is completely specified by its covariance function, or equivalently by its incremental variance

$$d^2(s, t) := \mathrm{E}[(X_s - X_t)^2].$$

Being an incremental standard deviation, d gives a metric on the parameter-space T

(by the Cauchy–Schwarz inequality). Let N_ε be the minimum number of balls of radius $\leqslant \varepsilon$ required to cover T. The necessary and sufficient condition for X to have continuous paths is (Fernique [**9**])

$$\int_{0+} \sqrt{\log N_\varepsilon}\, d\varepsilon < \infty$$

(sufficiency was proved earlier by R. M. Dudley in 1967). In the stationary case ($d(s,t) = \sigma(|s-t|)$, say) the Fernique criterion becomes

$$I(\bar{\sigma}) := \int_{0+} \bar{\sigma}(u)du/(u \log^{\frac{1}{2}}(1/u)) < \infty,$$

where $\bar{\sigma}$ is the non-decreasing rearrangement of σ (Jain and Marcus [**13**]). For a survey of earlier work, see Dudley [**5**]; in particular, the results above give in the Brownian case path-continuity and Lévy's modulus of continuity (*ibid.*, Ex. 2.2). Note also the dichotomy of Yu. K. Belyaev (another pupil of Kolmogorov's): Gaussian sample paths are either continuous, or unbounded on every interval (*ibid.*, §3.3). Remarkably, much of the above theory extends also to the non-Gaussian case (Fernique [**10**]).

For a subset A of a metric space, define $N_\varepsilon(A)$ as the number of balls of radius ε needed to cover A. Then $H_\varepsilon(A) := \log N_\varepsilon(A)$ is called the ε-entropy of A (Kolmogorov [1956g]; Kolmogorov and Tikhomirov [1959b]), or in the terminology of G. G. Lorentz [**23**], the metric entropy of A. This concept has proved very valuable in a number of contexts; for applications to approximation theory see Lorentz's survey, and for empirical processes, Dudley [**6**, Chapter 6].

6. *Other topics*

6.0. Markov processes. In his classic paper [1931a] Kolmogorov develops the theory of Markov processes, and in particular the Chapman–Kolmogorov equations. This subject is closely allied to the theory of Markov chains, discussed elsewhere; for a full modern treatment see the monograph of E. B. Dynkin [**8**], a pupil of ANK.

6.1. Statistics. Kolmogorov's interest in statistics is reflected in the title *Theory of probability and mathematical statistics* ('PS') of the second volume of his selected works. This contains papers on least squares, unbiased estimation and other topics; see [1931c, 1946b, 1947c, 1947f, 1950a].

6.2. Prediction and filtering. Following the work of 1938 by H. Wold on time series, Kolmogorov ([1939e, 1941a, b]) turned in 1938 to the prediction (or filtering) problem for stationary sequences. Similar work by Wiener, restricted during wartime, was published in 1949. A treatment of the 'Kolmogorov–Wiener filter' is given in Chapter XII of Doob [**4**]. This involves prediction given the infinite past; for the more difficult problem of prediction given a finite segment of the past, see Dym and McKean [**7**]. A recursive treatment of filtering was not developed till Kalman's work of 1960.

6.3. Branching processes and biological problems. Kolmogorov was deeply interested in biological problems; in particular he greatly admired the work of R. A. Fisher on genetics. We mention [1940g] on Mendel's laws (this led to a dispute with Lysenko; Kolmogorov's stand here took great courage in 1940); [1947d] with Dmitriev on branching processes (here, and independently in T. E. Harris's paper of the same year, the term branching process is introduced); [1937d] by Kolmogorov, Petrovskii and Piskunov ('KPP-Fisher') on the advancing-wave problem in genetics – this and Fisher's work each appeared in 1937, independently.

6.4. Self-similarity. In the study of self-similarity, initiated by B. B. Mandelbrot in the 1960s, the fractional Brownian motion process plays a distinguished role. It is interesting to note that this process was introduced by Kolmogorov in [1940c].

References

1. P. BILLINGSLEY, *Convergence of probability measures* (J. Wiley and Sons, 1968).
2. N. H. BINGHAM, 'The work of A. N. Kolmogorov on strong limit theorems', *Theory Probab. Appl.* 34 (1989) 152–169.
3. K. L. CHUNG, *A course in probability theory* (Academic Press, 1968).
4. J. L. DOOB, *Stochastic processes* (J. Wiley and Sons, 1953).
5. R. M. DUDLEY, 'Sample functions of Gaussian processes', *Ann. Probab.* 1 (1973) 66–103.
6. R. M. DUDLEY, 'A course in empirical processes', *École d'Été, Saint-Flour, XII.* Lecture Notes in Math. 1097 (Springer, Berlin, 1984), pp. 1–142.
7. H. DYM and H. P. MCKEAN, *Gaussian processes, function theory and the inverse spectral problem* (Academic Press, 1976).
8. E. B. DYNKIN, *Markov processes*, Volumes 1, 2 (Springer, Berlin, 1965).
9. X. FERNIQUE, 'Régularité des trajectoires des fonctions aléatoires gaussiennes', *École d'Été, Saint-Flour, IV.* Lecture Notes in Math. 480 (Springer, Berlin, 1975), pp. 1–91.
10. X. FERNIQUE, 'Régularité des trajectoires des fonctions aléatoires non gaussiennes', *École d'Été, Saint-Flour, XI.* Lecture Notes in Math. 976 (Springer, Berlin, 1983), pp. 1–74.
11. I. A. IBRAGIMOV and YU. V. LINNIK, *Independent and stationary sequences of random variables* (Wolters-Noordhoff, Groningen, 1971).
12. J. JACOD and A. N. SHIRYAYEV, *Limit theorems for stochastic processes* (Springer, Berlin, 1987).
13. N. C. JAIN and M. B. MARCUS, 'Sufficient conditions for the continuity of stationary Gaussian processes and applications to random series of functions', *Ann. Inst. Fourier* 24 (1974) 117–141.
14. M. KAC, *Statistical independence in probability, analysis and number theory*, Carus Math. Monographs 12 (J. Wiley and Sons, 1959).
15. M. KAC, 'Probability, number theory and statistical physics', *Selected papers* (ed. K. Baclawski and M. D. Donsker, MIT Press, 1979).
16. M. KAC, *Enigmas of chance, an autobiography* (Harper and Row, New York, 1985).
17. P. LÉVY, *Calcul des probabilitiés* (Gauthier-Villars, Paris, 1925).
18. P. LÉVY, *Théorie de l'addition des variables aléatoires* (Gauthier-Villars, Paris, 1937).
19. P. LÉVY, *Processus stochastiques et mouvement brownien* (Gauthier-Villars, Paris, 1948).
20. P. LÉVY, *Quelques aspects de la pensée d'un mathématicien* (A. Blanchard, Paris, 1970).
21. M. LOÈVE, *Probability theory* (2nd ed., Van Nostrand, 1963).
22. M. LOÈVE, 'Obituary: Paul Lévy, 1886–1971', *Ann. Probab.* 1 (1973) 1–18.
23. G. G. LORENTZ, 'Metric entropy and approximation', *Bull. Amer. Math. Soc.* 72 (1966) 903–937.
24. J. MARCINKIEWICZ, *Collected papers* (PWN, Warsaw, 1964).
25. V. V. PETROV, *Sums of independent random variables* (Springer, Berlin, 1975).
26. G. R. SHORACK and J. A. WELLNER, *Empirical processes with applications to statistics* (J. Wiley and Sons, 1986).
27. E. SLUTSKII, 'Qualche proposizione relativa alla teoria delle funzioni aleatorie', *Giorn. Ist. Ital. Attuari* 8 (1937) 183–199.

KOLMOGOROV'S CONTRIBUTION TO FOURIER SERIES

W. K. HAYMAN

Some of ANK's earliest work was on Fourier series, and as David Kendall has said in his introductory essay, it brought him instant fame.

Let f be a Lebesgue integrable function in $(0, 2\pi)$ and let

$$\tfrac{1}{2}a_0 + \sum_{1}^{\infty} (a_n \cos nx + b_n \sin nx) \tag{1}$$

be the Fourier series of f. In [1923a] ANK constructed a Fourier series that diverges almost everywhere, and in [1925h] he even constructed an everywhere divergent Fourier series.

Kolmogorov also obtained a number of supplementary conditions under which (1) converges. Thus in [1925g] with G. A. Seliverstov he showed that the condition

$$\sum (a_n^2 + b_n^2) \log n < \infty \tag{2}$$

is sufficient for the series (1) to converge almost everywhere. This remained the best result until Carleson [**1**] in 1966 proved that if $f \in L^2$, or equivalently if

$$\sum (a_n^2 + b_n^2) < \infty, \tag{3}$$

then (1) converges almost everywhere.

Carleson's theorem has been extended by Sjölin [**5**], who proved that

$$f \in L \log L \log\log L$$

is sufficient for almost everywhere convergence.

In the opposite direction Y. M. Chen [**2**] extended Kolmogorov's theorem slightly. Using Kolmogorov's technique he constructed a function in $L(\log\log L)^{1-\varepsilon}$ whose Fourier series diverges almost everywhere.

Such results lay far in the future in the nineteen-twenties. In [1927b] Kolmogorov and Menchov showed that if $W(n) = o(\log n)$ then there exist bounded orthogonal functions $\phi_n(x)$ for which

$$\sum (a_n^2 + b_n^2) W(n) < \infty,$$

while the series $\sum a_n \phi_n(x)$ diverges almost everywhere. This made it plausible to suppose that (2) might indeed be the right condition for a Fourier series to converge almost everywhere. In [1927b] it was also claimed that Kolmogorov had a proof that there exists a series

$$\sum a_n \cos(m_n x + \lambda_n) \tag{4}$$

that diverges almost everywhere, while (3) holds and the m_n are distinct integers. If the m_n form an increasing sequence then this would conflict with Carleson's theorem. At any rate it does not seem that Kolmogorov ever published a proof of this claim.

Another seminal paper of ANK was [1925d]. Suppose that

$$\tilde{f} = \sum_1^\infty (a_n \sin nx - b_n \cos nx)$$

is the conjugate series to f (which can be shown to be summable in a suitable sense). Kolmogorov showed that $\tilde{f}$ satisfies what is nowadays called a weak type (1, 1) inequality. Thus if $E(\lambda)$ is the set on which $|\tilde{f}| > \lambda$, then the measure $|E(\lambda)|$ of $E(\lambda)$ satisfies

$$|E(\lambda)| \leqslant \frac{C}{\lambda} \frac{1}{2\pi} \int_0^{2\pi} |f(\theta)| d\theta, \tag{5}$$

and, if $0 < p < 1$,

$$\left(\frac{1}{2\pi} \int_0^{2\pi} |\tilde{f}(\theta)|^p d\theta \right)^{\frac{1}{p}} \leqslant C(p) \frac{1}{2\pi} \int_0^{2\pi} |f(\theta)| d\theta. \tag{6}$$

The best values for the constants,

$$C = \left(\frac{1}{\pi}\int_0^\pi \left|\log\cot\left(\frac{1}{2}\theta\right)\right| d\theta\right)^{-1}$$

and

$$C(p) = \left(\frac{1}{\pi}\int_0^\pi (\sin\theta)^{-p}\, d\theta\right)^{\frac{1}{p}},$$

have only recently been found by B. Davis [3, 4].

I should also like to mention the following result in [1935d]. Suppose that $p \geqslant 1$ and that $c_n^{(p)}$ is the least upper bound of the Fourier remainder,

$$R_n(f,x) = f(x) - \tfrac{1}{2}a_0 - \sum_1^n (a_k \cos kx + b_k \sin kx)$$

for all f with a continuous $(p-1)$th derivative in $(0, 2\pi)$ which satisfies

$$|f^{(p-1)}(x) - f^{(p-1)}(y)| \leqslant |x-y|.$$

Then Kolmogorov proved that

$$c_n^{(p)} = \frac{4(\log n + O(1))}{\pi^2 n^p}$$

as $n \to \infty$. When p is odd he even found the exact value

$$c_n^{(p)} = \frac{1}{2\pi}\int_0^{2\pi} \left| \sum_{k=n+1}^{\infty} \frac{\sin kx}{k^p} \right| dx.$$

In a rather different direction lies the following theorem relating bounds for derivatives of functions on the real line $\mathbb{R}$. We write

$$M_k = \sup_{x\in\mathbb{R}} |f^{(k)}(x)|.$$

Then Kolmogorov announced in [1938o] and proved in [1939c] the following convexity theorem. If M_0 and M_n are finite then

$$M_k \leqslant C_{k,n} M_0^{1-k/n} M_n^{k/n} \; (0 < k < n),$$

with explicit values for the constants $C_{k,n}$.

Mention should also be made of the book on the theory of functions and functional analysis by Kolmogorov and Fomin ([1954l, 1960h] and four later editions), which is greatly treasured by analysts.

Kolmogorov's work on Fourier series was truly seminal. He proved that a Fourier series need not converge anywhere, and established conditions for almost everywhere convergence that were widely believed to be best possible, and were only sharpened 40 years later by Carleson. But possibly his greatest contribution is the work in [1925d] on conjugate series. His representation of the conjugate function and the proof of (5) and (6) led the way to the work on the Hilbert transform which lies at the heart of much of modern harmonic analysis.

References

1. L. Carleson, 'On convergence and growth of partial sums of Fourier series', *Acta Math.* 116 (1966) 135–157.
2. Y. M. Chen, 'An almost everywhere divergent Fourier series of the class $L(\log^+ \log^+ L)^{1-\varepsilon}$', *J. London Math. Soc.* 44 (1969) 643–654.
3. B. Davis, 'On the weak type (1, 1) inequality for conjugate functions', *Proc. Amer. Math. Soc.* 44 (1974) 307–311.
4. B. Davis, 'On Kolmogorov's inequalities $\|\tilde{f}\|_p \leqslant C_p\|f\|_1 (0 < p < 1)$', *Trans. Amer. Math. Soc.* 222 (1976) 179–192.
5. P. Sjölin, 'An inequality of Paley and convergence a.e. of Walsh–Fourier series', *Ark. Mat.* 7 (1969) 551–570.

KOLMOGOROV'S WORK IN LOGIC

J. M. E. Hyland

1. *Introduction*

Kolmogorov's two papers on logic are concerned almost entirely with intuitionism. Intuitionism is the philosophy and practice of mathematics that arose in the first instance from Brouwer's critique of classical mathematical reasoning. Brouwer held that mathematics consists of freely created mental constructions, and that truth depends (in some way) on knowledge, and so on proof. It follows, according to Brouwer, that pure existence proofs, which do not at least in principle exhibit a witness, should be regarded as illegitimate. Such non-constructive arguments typically depend on argument by contradiction: hence notoriously on this account, the law of the excluded middle is not valid. Kolmogorov wrote two papers on intuitionistic logic and mathematics, the first [1925e] published in 1925 and the second [1932d] in 1932 (though dated 15 January 1931).

2. *O printsipe 'tertium non datur'* (1925)

In this paper Kolmogorov appears to accept Brouwer's critique of the general applicability of the law of the excluded middle in mathematics, and addresses the question of why the use of illegitimate principles has not proved disastrous. The answer sketched would now be formalized as a relative consistency result. *K* describes a translation A^* of formulae A such that if A is provable using classical reasoning then A^* is provable using intuitionistic reasoning. Since the translation of a classical inconsistency is an intuitionistic one, this shows that if intuitionistic reasoning is consistent, then so is classical reasoning.

To present the formal details of his translation, Kolmogorov needed to provide some intuitionistically correct formal system of logic. He argues for the intuitive correctness of a number of axioms, and (the standard) rules of inference. He is not concerned with axiomatics as such, and simply mentions some questions about completeness in passing. However, he does to a considerable extent anticipate Heyting's axiomatization of the intuitionistic predicate calculus (see Heyting [5]). The system that results is minimal propositional calculus with implication and negation. Thus it is a system for intuitionistic logic without the principle that everything follows from a contradiction. This principle was recognized as problematic at the time, and more recently Dummett [1] felt the need to argue for it explicitly. Kolmogorov states baldly that it cannot have any intuitive foundation. This concern with the problems of negation is taken up again in the second paper.

In 1923 Brouwer showed in informal fashion that for any proposition A, the propositions $\neg A$ and $\neg\neg\neg A$ are equivalent. (As we would now say, the operation of negation provides a Galois connection.) Kolmogorov seems to have been the first to realize the general logical significance of this: in intuitionistic logic the double negation of any logical operator behaves as in classical logic. He proves Brouwer's result and other standard properties of double negation in his formal system, and thus establishes the basic properties of the 'double negation' translation. This translation is commonly attributed to Gödel in view of [2], where the translation is given for a full system of arithmetic. A restricted logic is sufficient for Kolmogorov's purposes. He gives a complete treatment for a component of propositional logic based on implication and negation, and sketches a treatment for the universal quantifier (the rules for which he takes to be part of the axioms of mathematics as opposed to logic).

In fact Kolmogorov wants to establish more than a (relative) consistency result. He regards the intuitionistic validity of a proposition of the form A^* as constituting the 'pseudo-truth' of the proposition A. So he claims that, notwithstanding the intuitionistic critique, classical mathematics can be preserved as the study of pseudo-truth. The result is that (contrary to the opinion of Brouwer) classical mathematics is conservative over constructive mathematics for finitary assertions. To establish all this, one should show that the non-logical, mathematical or set theoretic axioms of classical mathematics are at least pseudo-true. Kolmogorov simply states that this is the case, giving as example the laws for quantifiers (which we would take as part of logic). However, experience bears out his general impression. For a well-developed system of constructive mathematics such as that of Martin-Löf [10], it is quite straightforward to give an interpretation of classical type theory along the lines sketched by Kolmogorov. (Whether such systems faithfully reflect Brouwer's thought is another matter.)

3. *Zur Deutung der intuitionistischen Logik* (1932)

Kolmogorov's 1932 paper is not widely read. (It is written in German, while the earlier paper appears in English translation in van Heijenoort [4].) Unlike the first, the second paper is not written from the point of view of a committed intuitionist. Rather its first part contains an interpretation of intuitionistic logic that is intended to make perfectly good sense to a mathematician unimpressed by intuitionistic arguments. The interpretation is in terms of (unanalysed) notions of problem and solution. Kolmogorov takes intuitionistic logic in Heyting's axiomatization and argues that it can be identified with the classical logic of problems. This interpretation has a variety of descendants, described below.

In the second section of the 1932 paper the intuitionistic critique is itself examined. The argument is sketchy, but the general line of thought is clear. The intuitionist gives $\neg A$ positive force as the existential claim that there is a proof of a contradiction from A. Superficially the truth of existential assertions appears to depend on time; if $\exists x.A(x)$ means that one has constructed a witness w and a proof that $A(w)$ holds, then at some time it may be false, and later become true. Then, the paradoxical claim is that the whole point of intuitionism disappears because the only assertions whose negations have definite sense are the decidable ones, that is, those for which the law of the excluded middle does hold! Clearly this time dependent reading of existential propositions is not intended by Brouwer. Kolmogorov concludes that to make sense of Brouwer's thought, we need to consider a quite different kind of assertion, whose negation will have objective sense. This notion is more like that of a problem or task, whose subjective element is the solution. Thus Kolmogorov arrives at a vision of mathematics in which the solution of problems is considered as an independent task alongside the proof of theoretical assertions. Kolmogorov's paradox of intuitionism is not compelling. A modern account of intuitionism would present a time-independent reading: roughly, 'A' is equivalent to 'A is provable' rather than 'I have a proof of A'. However, Kolmogorov's apparent (pre-recursion theory) vision, of a kind of constructive problem mathematics in harness with a (presumably) classical theoretical mathematics, is appealing in itself. In fact Kolmogorov does not give any account of the (logic of the) theoretical assertions, to which he refers.

It is hard to be sure of Kolmogorov's position in this very short section. For example, in a footnote he raises the 'new question' of which laws of logic hold good for propositions whose negation has no sense. If this were just a case of finding an axiom system, he had already raised this question (for the implicational fragment of classical logic) in his 1925 paper. What did he have in mind?

Kolmogorov's paper was written independently of Heyting's similar work on the interpretation of intuitionistic logic [**6**]. Heyting analysed a proposition as an expectation to find some condition fulfilled; he took himself to be making explicit the proper intuitionist's conception of the meaning of propositions in terms of constructions or proofs, implicit in the writings of Brouwer. Kolmogorov on the other hand is not concerned to elucidate the intuitionistic critique: he simply provides an interpretation of the logic. Over the years we have come rightly or wrongly to regard Kolmogorov's interpretation and Heyting's explanation as different readings of the same idea. This view is clearly presented in Martin-Löf [**10**]. Martin-Löf's own position, which descends from Kolmogorov via the (Curry–Howard) 'Propositions as Types' interpretation, involves identifying propositions as problems; an independent notion of abstract assertion is allowed no clear sense.

A variety of formal interpretations (Kleene [**7**], Gödel [**3**], Kreisel [**8, 9**]) in the style of Heyting and Kolmogorov have played a major role in the study of the metamathematics of intuitionistic systems. A recent book (Troelstra and van Dalen [**11**]) calls such formal interpretations 'BHK-interpretations', so that the main early protagonists (Brouwer, Heyting and Kolmogorov) all get credit. Kreisel [**9**] also contains an illuminating and non-dogmatic discussion of the foundational issues involved in intuitionism.

4. *Conclusion*

There is a shift in emphasis between Kolmogorov's two papers on logic. The first provides an interpretation of classical mathematics within intuitionistic mathematics. Clearly here, Kolmogorov finds Brouwer's intuitionism a great deal more congenial than Hilbert's formalism. The second paper can be read somewhat less surely in the reverse direction; it provides an interpretation of intuitionistic logic within classical mathematics. While it does not arrive at a definite philosophy of mathematics, it contains a clear attempt to preserve the intuitionistic insights in some form.

Both papers treat what have proved to be fundamental ideas about intuitionistic mathematics. They were written over sixty years ago in what is effectively the pre-history of mathematical logic. Though the style is of that time, the ideas still seem fresh and the attitude is remarkably modern.

References

1. M. A. E. DUMMETT, *Elements of intuitionism* (Clarendon Press, Oxford, 1977).
2. K. GÖDEL, 'Zur intuitionistischen Arithmetik und Zahlentheorie', *Ergebnisses eines math. Koll.* 4 (1933) 34–38.
3. K. GÖDEL, 'Über eine bisher noch nicht benutze Erweiterung des finiten Standpunktes', *Dialecta* 12 (1958) 280–287.
4. J. VAN HEIJENOORT (ed.), *From Frege to Godel: a source book in mathematical logic, 1879–1931* (Harvard University Press, 1967).
5. A. HEYTING, 'Die formalen Regeln der intuitionistischen Logik', *Sitzenberichte Preuss. Akad. Wiss., Phys-Math. Kl.* (1930) 42–56.
6. A. HEYTING, 'Die intuitionistische Grundlegung der Mathematik', *Erkenntnis* 2 (1931) 106–115.
7. S. C. KLEENE, 'On the interpretation of intuitionistic number theory', *J. Symb. Logic* 10 (1945) 109–124.
8. G. KREISEL, 'Interpretation of analysis by means of constructive functionals of finite type', *Constructivity in mathematics* (ed. A. Heyting, North-Holland, Amsterdam, 1959).
9. G. KREISEL, 'Foundations of intuitionistic logic', *Logic, methodology and philosophy of science* (ed. E. Nagel, P. Suppes and A. Tarski, Stanford University Press, 1962), pp. 198–210.
10. P. MARTIN-LÖF, Intuitionistic type theory (Bibliopolis Napoli, 1984).
11. A. S. TROELSTRA and D. VAN DALEN, *Constructivism in mathematics, an introduction*, Volumes I and II (North-Holland, Amsterdam, 1988).

SUPERPOSITIONS, METRIC ENTROPY, COMPLEXITY OF FUNCTIONS, WIDTHS

G. G. LORENTZ

I will review the work of Kolmogorov on these subjects; all of it except for [1936a] was achieved in 1956–63. Sometimes ANK wrote only one paper, a seminal one, leaving the field wide open (§4). At other times, the initial result of ANK was a lonely peak, around which the ensuing work clustered (§1). In each section I will try to discuss also the later work that was a logical development of his. First, however, I would like to give some of my personal thoughts and reminiscences about ANK.

One is astonished to see how much of ANK's best work looks even better now than when it first appeared. An example is his 1923 function $f \in L_1$ with almost everywhere divergent Fourier series. At the time this was thought to be a first step towards proving the Luzin conjecture, which postulated the existence of continuous functions with a.e. divergent Fourier series. And we know now that the series converge a.e. even for functions $f \in L_p, p > 1$.

One could ask, was ANK a pure or an applied mathematician? Is it possible to make this distinction? Without attempting to answer the latter question, let us assume that the answer is yes. Then, what astonishes one is how much of ANK's work (including probability) could, with some justification, be called applied mathematics – perhaps as much as two thirds. Yet generally ANK is considered a pure mathematician. My explanation of this is as follows. What had been applied mathematics before ANK's investigation, through his vision became a part of science where applied and pure mathematics intermingle, become indistinguishable.

The only time I have spoken with ANK personally was at a party given by Professor Fichtenholz in Leningrad in the late 1930s. I was a young assistant professor then. As I recall, ANK made a mathematical joke about the position of a

The article by G. G. Lorentz has been partially supported by Texas ARP.

rower on the Moscow river (assuming him to be I. M. Vinogradov). Wanting to make a joke of my own, I began 'There are two points on the line...'. ANK gently interrupted: 'Where else could they be?' This reduced me to silence.

At the same meeting, ANK said that he was absolutely sure that Euler's constant γ is transcendental, but that the problem is 'naked' in the sense that there are no approaches to its solution. Perhaps he tried to solve it himself? He also described what in his opinion was a good doctoral dissertation, meaning, of course, a Russian one. According to ANK, a good dissertation 'must astonish'. (It remained unclear who must be astonished. ANK himself?) As an example he mentioned Tikhonov's dissertation on partial differential equations.

1. *Superpositions*

The solution of the 13th problem of Hilbert is the most spectacular of ANK's achievements discussed in this article. At the second International Congress of Mathematicians at Paris in 1900, Hilbert formulated this problem in the following way.

(I) *Prove that the equation of seventh degree* $x^7+ax^3+bx^2+cx+1=0$ *is not solvable with the help of any continuous functions of only two variables.*

If we ignore the algebraic part of the problem, this can be interpreted to mean:

(II) *Prove that there are continuous functions of three variables not representable by continuous functions of two variables.*

The means that we are allowed to use are superpositions of functions. For example

$$f(x,y,z)=F(g(x,y),h(\phi(x),\psi(x,z))) \tag{1.1}$$

is a superposition of functions of one and of two variables. A negative answer to (II) will, of course, give a negative answer to (I). In the years before ANK's paper [1957b] there were some unsuccessful attempts to prove (II). After preliminary work, partly with Arnold, ANK straightforwardly *disproved* (II). He established:

THEOREM 1.1. *All continuous functions of two or more variables are superpositions of functions of one variable, and of the special function* $s(x,y)=x+y$ *of two variables. More precisely, each continuous function* f *on* I^n, *where* $I=[0,1]$, *has a representation*

$$f(x_1,\dots,x_n)=\sum_{j=1}^{2n+1} g_j\left(\sum_{i=1}^{n}\phi_{i,j}(x_i)\right) \tag{1.2}$$

where the $\phi_{i,j}$ *are fixed continuous functions on* I, *while the* $g_j\in C(\mathbb{R})$ *are continuous functions that depend on* f.

Fairly soon, by efforts of Fridman in 1972, of Sprecher in 1965 and of the present author, this has been improved to:

THEOREM 1.2. *Each function* $f\in C(I^n)$ *has a representation*

$$f(x_1,\dots,x_n)=\sum_{j=1}^{2n+1} g\left(\sum_{i=1}^{n}\lambda_i\phi_j(x_i)\right) \tag{1.3}$$

where the ϕ_j *are fixed strictly increasing functions from* Lip 1 *on* I, *the* λ_i *are fixed algebraically independent constants, and only* $g\in C(\mathbb{R})$ *depends on* f.

In (1975) Kahane simplified the proof of (1.3) using category arguments. The following is a geometric formulation of Theorem 1.2.

THEOREM 1.3. *There exists a homeomorphic embedding h of I^n into $\mathbb{R}^{2n+1}$ so that on the image $B = h(I^n)$ of I^n, each continuous function f is of the form*

$$f(y_1, \dots, y_{2n+1}) = \sum_{j=1}^{2n+1} g(y_j) \tag{1.4}$$

for some $g \in C(\mathbb{R})$. The embedding h is of the special form

$$y_j = \sum_{i=1}^{n} \lambda_i \phi_j(x_i), \; j = 1, \dots, 2n+1.$$

We can compare this with the classical topological theorem (due to Nöbeling) according to which each n-dimensional separable metric space can be embedded into $\mathbb{R}^{2n+1}$. We see that ANK's theorem is a refined version of a special case of this.

The number $2n+1$ of (1.2) and (1.3) appears also in a fine combinatorial lemma of ANK about coverings of $\mathbb{R}^n$ by $2n+1$ systems of disjoint cubes. It enables one to derive a 'baby form' of (1.2) and appears in all known proofs of ANK's theorem.

After Theorem 1.2, all work on ANK's theorem has been 'negative', that is, it consists of proofs that certain aspects of this theorem cannot be improved.

First, do there exist at all genuine continuous functions of several variables apart from the sum $x+y$? Theorems 1.1 and 1.2 deny this. Thus the product xy reduces to the sum and to functions of one variable: $xy = \exp(\log x + \log y)$. However, Hilbert's conjecture was based on a sound idea, that not all in a sense bad functions (functions of many variables) can be represented by good functions (functions of few variables). Indeed Vitushkin **[12]**, even before ANK's theorem, proved 'the Fundamental Theorem of Differential Calculus': there exist differentiable functions of arbitrarily many variables. More precisely, we have:

THEOREM 1.4. *For $r \geqslant 1$ and $n \geqslant 2$ there exist r-times continuously differentiable functions of n variables, not representable by r-times continuously differentiable functions of fewer variables.*

A simple proof of this by Kolmogorov and Tikhomirov [1959b] uses the estimate (2.3) of the metric entropy of the unit ball of the Sobolev space and the fact that this entropy is not essentially increased under superpositions. In other words, the entropy of the set of bad functions is larger than that of a set of good functions. It is necessary here to measure the quality of a function by means of the ratio r/n.

Next question: is it possible to replace the *Lipschitz functions* ϕ in (1.3) and (1.2) by *continuously differentiable* functions? No, in this respect Theorem 1.2 is the best possible.

We call an expression of the form

$$\sum_{j=1}^{N} p_j(x_1, \dots, x_n)\, g_j(\phi_{j,1}, \dots, \phi_{j,n-1}) \tag{1.5}$$

with $\phi_{j,k} = \phi_{j,k}(x_1, \dots, x_n)$, a *linear superposition*. For example, (1.2) and (1.3) are linear superpositions. Now Kaufman **[3]** (and the present author) proved:

THEOREM 1.5. *Not all functions $f \in C(I^n)$ are representable by means of linear superpositions* (1.5) *with fixed continuous functions p_j and $\phi_{j,k} \in C(I^n)$ and arbitrary $g_j \in C(I^{n-1})$, because they form a set of first category in $C(I^n)$.*

Earlier, Vitushkin and Henkin [**14**] proved this when the $g_j \in C(I)$ in (1.5) depend only on one variable. Then the functions (1.5) form a *nowhere dense* set in $C(I^n)$. It is not known whether this holds for the general case in (1.5).

Leaving aside now the special properties which the functions $\phi_{i,j}$ of (3.2) may have, we ask probably the most important question: what is the minimal number of terms in (1.2)? And ANK's number $2n+1$ proves to be the best possible. This was shown only in 1985 by Sternfeld. Let X, $Y_j, j = 1, \ldots, N$ be compact metric spaces. A family $F = \{\phi_j\}_{j=1}^N$ of continuous functions which map X into Y_j is said to be a *basic family for X* if each $f \in C(X)$ admits a representation

$$f(x) = \sum_{j=1}^{N} g_j(\phi_j(x)), \quad x \in X, \tag{1.6}$$

for some $g_j \in C(Y_j)$, $j = 1, \ldots, N$. Thus, in the case $X = I^n$, $N = 2n+1$, $Y_j = I$, $j = 1, \ldots, N$, Theorem 1.1 implies that there exists a basic family $\{\phi_j\}_1^N \subset C(X)$ and even one of the special form $\phi_j(x_1, \ldots, x_n) = \sum_{i=1}^n \phi_{i,j}(x_i), j = 1, \ldots, N$. Using topological as well as combinatorial arguments, Sternfeld [**10**] established:

THEOREM 1.6. *If X is a compact metric space that has topological dimension* $\dim X = n$, *then each basic family $F \subset C(X)$ contains at least $2n+1$ functions.*

As a corollary, the number of terms in (1.2) cannot be reduced even for much more general representations

$$f(x_1, \ldots, x_n) = \sum_{j=1}^{2n+1} g_j(\phi_j(x_1, \ldots, x_n)).$$

2. *Metric entropy*

This notion is different from the 'probabilistic entropy', treated in the article of W. Parry; it is a realization of ideas of Shannon [**8**], in application to metric spaces X. A compact set $A \subset X$ can be approximately described by an economical ε-net ($\varepsilon > 0$) of points $(x_i)_{i=1}^N$ with the property that the distance of each point $x \in A$ to at least one of the x_i is $< \varepsilon$. Then A is covered by N balls of centres x_i and radii ε. This leads us to define $N_\varepsilon(A)$ to be the smallest N with the property just formulated. Then

$$H_\varepsilon(A) := \log N_\varepsilon(A) \tag{2.1}$$

is called the *metric entropy* of A. Normally, $H_\varepsilon(A) \to \infty$ for $\varepsilon \to 0$, and the asymptotic behaviour of this function describes the 'massiveness' of A. There is a dual notion of the *metric capacity* $C_\varepsilon(A)$ of A. This is $C_\varepsilon(A) := \log M_\varepsilon(A)$, where $M_\varepsilon(A)$ is the minimum of the cardinal numbers M of sets of ε-distinguishable points $(y_i)_1^M$ of A, that is, points with $\operatorname{dist}(y_i, y_j) > \varepsilon, i \neq j$.

The general theory of metric entropy and capacity is not rich. The main interest lies in the asymptotic determination of entropy for concrete compact sets in function spaces. For this purpose, one invariably uses the inequality

$$C_{2\varepsilon}(A) \leqslant H_\varepsilon(A). \tag{2.2}$$

If one can estimate $C_\varepsilon(A)$ well from below, and $H_\varepsilon(A)$ from above, then a good estimate of both will result.

In [1956g] ANK gave these definitions and some examples, followed by his joint paper (1959b) with Tikhomirov, with a rich collection of sets of functions with a calculation of their entropy for the uniform metric. The two main results (the second is by Vitushkin [13]) are:

$$H_\varepsilon(B_q^n) \approx C_\varepsilon(B_q^n) \approx (1/\varepsilon)^{1/q}, \quad q = p+\alpha, \tag{2.3}$$

where B_q^n is the set of functions on a compact region in $\mathbb{R}^n$, with all partial derivatives of order $\leqslant p$ satisfying the α-Lipschitz condition, $0 < \alpha \leqslant 1$. Moreover,

$$H_\varepsilon(A) \text{ and } C_\varepsilon(A) = C\log^{n+1}\frac{1}{\varepsilon} + O\left(\log^n\frac{1}{\varepsilon}\log\log\frac{1}{\varepsilon}\right), \tag{2.4}$$

for the set A of uniformly bounded analytic functions of n variables on a multidisc, with the uniform norm on a smaller concentric multidisc; the constant C depends on the radii of the multidiscs.

If $T: X \to Y$ is a *compact* linear operator from X to Y, one defines the entropy of T by $H_\varepsilon(T) := H_\varepsilon(B')$, where $B' = T(B_X)$ stands for the image of the unit ball of X under T. This is an important tool in the investigation of compact operators and the geometry of Banach spaces. Work of Carl, Pietsch, Triebel, V. Milman and others attest to this. The eigenvalues $\lambda_n(T)$ of T (taken in decreasing order of $|\lambda_n(T)|$) are important in this theory, as also are the so-called s-numbers. To this category belong the *entropy widths* $\varepsilon_n(T)$ of T. For $n = 1, 2, \ldots, \varepsilon_n(T) = \varepsilon$ is obtained by solving the equation $H_\varepsilon(T) = n$. An example is the inequality of Carl:

$$|\lambda_n(T)| \leqslant \sqrt{2}\varepsilon_n(T), \tag{2.5}$$

which holds for a compact operator T mapping a Banach space into itself.

It is perhaps worthwhile to discuss a general approach to the estimates such as (2.3) and (2.4). What is desirable is to have an approach which will: (a) work for arbitrary Banach spaces X; (b) depend only on the geometry of the set $A \subset X$ and not on special properties of elements of A (such as analyticity); (c) be good enough to yield estimates such as (2.3) and (2.4), even with the remainder term.

This proves to be possible (Lorentz [4]) for the so-called full approximation sets (or balls in approximation spaces, in the terminology of Butzer and Scherer – Levin and Tikhomirov apply these ideas to more general sets), which are defined by two sequences $(d_\iota)_1^\infty$, $d_\iota > 0$ and $\{\phi_\iota\}_1^\infty$, the latter spanning X. Then a full approximation set A is given by $A = \{f \in X : E_n(f) \leqslant d_n,\ n = 1, 2, \ldots\}$. The estimates of $H_\varepsilon(A)$ and $C_\varepsilon(A)$ are obtained by means of the theorem of Brunn and Minkowski about mixed volumes and a variant of another theorem of Minkowski, which counts the number of lattice points in a convex body in $\mathbb{R}^n$.

There are numerous applications of entropy. As examples we mention applications to: non-linear approximation (Vitushkin [13]); as topological entropy to ergodic theory (Dinaberg, 1970; Bowen, 1971); to sound transcription and the coding of signals (Buslaev and Vitushkin, 1975); to almost everywhere convergence of sequences of linear operators (Bourgain, 1988).

3. *Complexity of functions*

ANK's approach here is based on the number of bits of the memory of a computer that are needed in order to approximate a function f with an error $< \varepsilon$. In this theory, the approximations, for each f, are constructed 'from scratch', without the use of other ('known') functions.

A Boolean function F is any map of $\{0,1\}^n$ into $\{0,1\}^m$, that is, a map which assigns to each sequence $X: x_1, \dots, x_n$ of 0s and 1s a similar sequence $Y: y_1, \dots, y_m$ of length m. One can represent F as a superposition of N operations $1-x$ and $y \cup z$ on terms $x, y, z = 0, 1$. The complexity $K(F)$ of F is the number

$$K(F) = n + m + \min N. \tag{3.1}$$

Let $f \in C[0,1]$, $0 \leqslant f(x) \leqslant 1$, and $\varepsilon > 0$ be given. We select n and m so that $|f(x) - f(x')| < \frac{1}{2}\varepsilon$ for $|x - x'| < 2^{-n}$, and that $2^{-m} < \frac{1}{2}\varepsilon$. Then a Boolean function $F: x_1, \dots, x_n \to y_1, \dots, y_m$ is said to approximate f with error $< \varepsilon$ if $|x - \sum_{i=1}^{n} 2^{-i} x_i| \leqslant 2^{-n}$ implies $|f(x) - \sum_{j=1}^{m} 2^{-j} y_j| \leqslant \varepsilon$. (This means: the knowledge of F allows one to compute $f(x)$ with error $< \varepsilon$ at each x.)

The complexity $K(f, \varepsilon)$ of f is a function of $\varepsilon > 0$ defined by

$$K(f, \varepsilon) = \min\{K(F): F \text{ approximates } f \text{ with error } < \varepsilon\}. \tag{3.2}$$

For a set of functions $A \subset C[-1,1]$ we put $K(A, \varepsilon) = \sup_{f \in A} K(f, \varepsilon)$.

The short paper [1963c] of ANK on this subject does not contain proofs. Much later, they have been supplied by Asarin [**1**].

The ε-complexity $K(A, \varepsilon)$ of a class $A \subset C[-1,1]$ can be computed if A is a unit ball in a Sobolev space, or a set of bounded analytic functions. In these examples, $K(a, \varepsilon)$ proves to be very close to the capacity $C_\varepsilon(A)$ of §2. In the first case they are different by a factor of order $\log(1/\varepsilon)$.

For the complexity of individual functions there are no 'inverse theorems'. Thus a slow increase of $K(f, \varepsilon) \to \infty$ for $\varepsilon \to 0$ does not imply that f is smooth. As an example ANK gives van der Waerden's nowhere differentiable function f for which $K(f, \varepsilon)$ is smaller than the typical complexity of analytic functions.

Why has ANK's theory of complexity of individual functions never been further developed? One reason might be that ANK is no longer with us. Another reason might be that in today's computing one rarely starts 'from scratch'. Instead, one can use many other helpful functions, or use software. This explains the existence of attempts to estimate complexity using different ideas, such as the algebraic computational complexity. See the expository article [**9**] of Strassen. One of the first and perhaps the most striking results of this theory (Schönhage and Strassen [**7**]) asserts that multiplication of two large numbers with N dyadic digits each can be completed in $O(N \log N \log\log N)$ steps.

4. *Widths*

This subject was introduced in the short seminal paper of ANK [1936a], the only one that he wrote on the question. There is at present an enormous literature on this subject. See the excellent book of Pinkus [**6**].

If X_n is an n-dimensional subspace of a Banach space X, then for $f \in X$, $E_n(f) := \inf_{g \in X_n} \|f - g\|$ is the error of approximation of f by X_n. For a subset $K \subset X$, we measure $E_n(K)$, the error of approximation of K, by the worst approximable elements f of K:

$$E_n(K) := \sup_{f \in K} \inf_{g \in X_n} \|f-g\|. \tag{4.1}$$

Finally, for a given n, we look for an n-dimensional subspace X_n that approximates K best. We obtain *the* nth *Kolmogorov width of* K:

$$d_n(K,X) := d_n(K) := \inf_{X_n \subset X} \sup_{f \in K} \inf_{g \in X_n} \|f-g\|. \tag{4.2}$$

Subspaces X_n (if they exist) which realize this infimum, are called *the optimal subspaces for* K *in* X. The problem is to determine $d_n(K)$ (exactly or asymptotically) for some natural sets K. Many beautiful early results are due to Tikhomirov [**11**]; among other things, he recognized the significance of Borsuk's antipodal theorem in these questions. The most difficult problem, solved only recently (and not covered adequately in the book of Pinkus), concerns the nth widths of the unit ball B_p^r of the Sobolev space $W_p^r[a,b]$ (p refers here to the L_p metric) in $L_q[a,b]$. The point is that for $r = 1, 2, \ldots$ and $2 \leqslant p \leqslant q \leqslant \infty$ the optimal subspaces are the classical polynomial subspaces, and the $d_n(B_p^r, L_q)$ can be estimated fairly easily. For the remaining pairs p, q these natural subspaces are far from being optimal, and the true values of the $d_n(B_p, L_q)$ are much smaller than their natural upper estimates. The optimal spaces are then described only indirectly. Probabilistic methods were used by Kashin [**2**] to find the asymptotic behaviour of the $d_n(B_p^r)$ in these cases. A simpler proof by Gluskin in 1984 uses the P. Lévy–E. Schmidt isoperimetric theorem about measures of subsets of the Euclidean sphere of high dimension. This was further simplified by Makovoz [**5**].

The results of ANK's paper [1936a] are characteristic for the width theory in Hilbert spaces. ANK studies the class $B_2^r = \{f : \|f^{(r)}\|_2 \leqslant 1\}$ in $L_2[0,1]$ and proves that

$$d_n(B_2^r, L_2) = \lambda_n^{1/2}, \quad n \geqslant r, \tag{4.3}$$

where λ_n are the eigenvalues of the problem

$$\lambda(-1)^r y^{(2r)} - y = 0, \tag{4.4}$$

with the side conditions $y^{(k)}(x) = 0$, $k = r, \ldots, 2r-1$ for $x = 0$, $x = 1$. Besides the 'classical' optimal subspaces spanned by the eigenfunctions of (4.4), there are also optimal subspaces of splines. This fact appears also in other similar situations.

References

1. E. A. Asarin, 'On the complexity of uniform approximations of continuous functions', *Uspekhi Mat. Nauk* 39 no. 3 (237) (1984) 157–169.
2. B. S. Kashin, 'Diameters of some finite-dimensional sets and classes of smooth functions', *Izv. Akad. Nauk SSSR* 41 (1977) 334–351.
3. R. Kaufman, 'Linear superpositions of smooth functions', *Proc. Amer. Math. Soc.* 46 (1974) 360–362.
4. G. G. Lorentz, 'Metric entropy and approximation', *Bull. Amer. Math. Soc.* 72 (1966) 903–937.
5. Y. Makovoz. 'A simple proof of an inequality in the theory of n-widths', *Constructive theory of functions* (Bulgarian Acad. Sci., Sofia, 1988), pp. 305–308.
6. A. Pinkus, n-*widths in approximation theory* (Springer, Berlin, 1985).
7. A. Schönhage and V. Strassen, 'Schnelle Multiplikation grosser Zahlen', *Computing* 7 (1971) 281–292.
8. C. E. Shannon, *A mathematical theory of communication* (University of Illinois Press, Urbana, 1949).

9. V. Strassen, 'Algebraische Berechnungskomplexität', *Perspectives in mathematics* (Birkhäuser, Basel, 1984), pp. 509–550.
10. Y. Sternfeld, 'Dimension, superposition of functions on separation of points, in compact metric spaces', *Israel Math. J.* 49 (1985) 13–53.
11. V. M. Tikhomirov, 'Widths of sets in function spaces and the theory of best approximations', *Uspekhi Mat. Nauk.* 15 no. 3 (93) (1960) 81–120.
12. A. G. Vitushkin, 'On the 13th problem of Hilbert', *Dokl. Akad. Nauk SSSR* 95 (1954) 701–704.
13. A. G. Vitushkin, *Theory of the transmission and processing of information* (Pergamon Press, New York, 1961; Russian edition: 1959).
14. A. G. Vitushkin and G. M. Henkin, 'Linear superpositions of functions', *Uspekhi Mat. Nauk* 22 (1967) 77–124.

KAM-THEORY

H. K. Moffatt

In 1953 and 1954 Kolmogorov wrote two papers [1953c 1954a] on the general theory of dynamical systems, with important applications to Hamiltonian mechanics. Both papers were precisely four pages in length, the limit permitted by *Doklady Akad. Nauk SSSR* at that time. Their influence on the subsequent development of the subject has however been out of all proportion to their length; indeed the second paper, whose title may be translated 'On the preservation of quasi-periodic orbits under a small change of Hamiltonian' contains the essence of what has subsequently come to be known as KAM-theory (after *K*olmogorov, *A*rnold and *M*oser). Kolmogorov stated the first critical theorems in this field and outlined the essential ingredients in their proof; it was left to V. I. Arnold [1] and J. Moser [2] to complete the proofs and to extend somewhat the circumstances to which Kolmogorov's theorems apply. KAM-theory lies at the heart of recent new understanding of the phenomenon of chaos in Hamiltonian systems (see, for example, Percival [3] in the proceedings of the Royal Society Discussion Meeting on Dynamical Chaos held in February 1987), and Kolmogorov's contribution in 1954 may be seen, with the benefit of hindsight, as providing the most important breakthrough in this subject since the fundamental difficulties were first recognized by Poincaré [4] in 1892. I say 'with the benefit of hindsight' because it was not until the development of the high-speed computers of the 1970s and 1980s that the full significance of KAM-theory could be properly appreciated. There were in fact rather few citations of Kolmogorov's papers on this subject up to about 1970; and since then the papers of Arnold and Moser, being more accessible to English-speaking readers, are those that are most widely known. There can be no doubt however that Kolmogorov was the ultimate source of inspiration for these new developments.

It is worth noting that Kolmogorov presented an account of this work, referring to both of the papers mentioned above, in a lecture at the International Congress of Mathematicians held in Amsterdam 2–9 September 1954; this lecture, 'General theory of dynamical systems and classical mechanics' appeared (in Russian) in Vol. 1 of the Proceedings of the Congress, published in 1957. Kolmogorov's famous paper 'On the preservation of quasi-periodic orbits...' was 'received' by *Doklady* on 31 August 1954 (and it appeared before the year was out!). One may surmise that it was partly the stimulus of preparing an important invited lecture to the International Congress that promoted the breakthrough for which Kolmogorov had already prepared the ground in his earlier (1953) paper.

What, then, was the nature of this breakthrough? Kolmogorov considers an autonomous Hamiltonian system with Hamiltonian $H(q_\alpha, p_\alpha, \theta)$ where α runs from 1 to 5 (the number of degrees of freedom), and θ is a perturbation parameter; he assumes 'for simplicity' that H is an analytic function of its arguments, although it is clear that he recognises that this assumption is stronger than absolutely necessary. After some preliminary statement of notation† he states his Theorem 1 (referred to later by Arnold as *Kolmogorov's theorem* – see, for example, V. I. Arnold *Mathematical methods of classical mechanics*, Springer-Verlag, 1978), which in view of its great historic interest we state here in full (in translation from the Russian original):

'THEOREM 1. Let

$$H(q, r, 0) = m + \sum_\alpha \lambda_\alpha p_\alpha + \frac{1}{2} \sum_{\alpha\beta} \Phi_{\alpha\beta}(q) p_\alpha p_\beta + O(|p|^3) \tag{2}$$

where m and λ_α are constants such that for suitable constants $c > 0$ and $\eta > 0$ the inequality

$$(n, \lambda) \geqslant \frac{c}{|n|^\eta} \tag{3}$$

is satisfied for all integer vectors n. Furthermore, let the determinant formed from the mean values

$$\phi_{\alpha\beta}(0) = \frac{1}{(2\pi)^s} \int_0^{2\pi} \dots \int_0^{2\pi} \Phi_{\alpha\beta}(q)\, dq_1 \dots dq_s$$

of the functions

$$\Phi_{\alpha\beta}(q) = \frac{\partial^2}{\partial p_\alpha \partial p_\beta} H(q, 0, 0)$$

be non-zero:

$$|\phi_{\alpha\beta}(0)| \neq 0. \tag{4}$$

Then there exist analytic functions $F_\alpha(Q, R, \theta)$ and $G_\alpha(Q, P, \theta)$ defined for all sufficiently small θ and for all points (Q, P) in some neighbourhood V of the set T_0, such that the associated contact transformation

$$q_\alpha = Q_\alpha + \theta F_\alpha(Q, P, \theta), \quad p_\alpha = P_\alpha + \theta G_\alpha(Q, P, \theta)$$

of V into $V' \subseteq G$ reduces H to the form

$$H = M(\theta) + \sum_\alpha \lambda_\alpha P_\alpha + O(|P|^2) \tag{5}$$

($M(\theta)$ does not depend on Q and P).'

Recognizing that the import of this theorem may be lost on the inexpert reader, Kolmogorov immediately provides a vital word of explication which again is worth quoting in full:

'It is easy to understand the importance of Theorem 1 for mechanics. It shows that the s-parameter family of quasi-periodic motions

$$q_\alpha = \lambda_\alpha t + q_\alpha^{(0)}, \quad p_\alpha = 0$$

existing at $\theta = 0$ cannot disappear under conditions (2) and (3) as a result of a small change in the Hamiltonian H: there is merely a shift of the s-

† Vector notation is used, for example, $p = (p_\alpha)$ with scalar product $(p, q) = \sum_\alpha p_\alpha q_\alpha$ and $|p|^2 = (p, p)$. The space G is the product of an s-dimensional torus T and a domain S of $\mathbb{R}^s$. It is assumed that $p = 0$ is contained in S; T_0 is the set of points in G for which $p = 0$. The Theorem 1 as quoted above contains an obvious misprint which is faithfully transcribed from the original Russian version.

dimensional torus T_0 which is covered by the trajectories of these motions into the torus $P = 0$, which remains covered with trajectories of quasi-periodic motions with the same set of frequencies $\lambda_1, \ldots, \lambda_s$.'

This style is not untypical of Kolmogorov: a theorem stated in full formality and at some length, followed by an informal and illuminating indication of its real meaning. This is further followed, not by a proof of the theorem (although it is hard to believe that Kolmogorov would have stated it as a theorem if he had not been 100% confident of its provability) but by a summary of the procedure by which the contact transformation $(q, p) \to (Q, P)$ may be constructed, in the course of which discussion, the need for and meaning of the conditions (3) and (4) of the theorem is made abundantly clear. For a system with two degrees of freedom, the condition (3) takes the form

$$n_1 \lambda_1 + n_2 \lambda_2 \geqslant \frac{c}{|n|^\eta} \qquad (*)$$

which, with n_1 and n_2 integers, means that the frequency ratio λ_1/λ_2 must be 'sufficiently irrational', a condition which appears also in the earlier (1953) paper. If resonances occur through vanishing of $n_1 \lambda_1 + n_2 \lambda_2$ (for any n_1, n_2) then Kolmogorov's procedure fails, just as Poincaré's attempt to analyse non-integrable systems by perturbation analysis had failed some 60 years before. But Kolmogorov's recognition of the need for a condition of the type $(*)$ was the crucial flash of insight that enabled vital progress to be made.

The torus $P = 0$ that survives the perturbation in the above theorem is of course what later came to be known as the KAM-torus; and Kolmogorov argued further in his (1954) paper that 'for small values of θ the displaced tori obtained in accordance with Theorem 1 fill the greater part of the region G', a statement which he then refined to a Theorem 2 (not a conjecture!) relating to the Lebesgue measure of the set of quasi-periodic orbits that survive the perturbation. And all this in four pages! The degree of crystallization of thought in these four pages is truly remarkable and can rarely have been surpassed.

References

1. V. I. Arnold, 'Small denominators II, Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally-periodic motions under a small perturbation of the Hamiltonian', *Russian Math. Surveys* 18 (1963) no. 5, 9–36 (translated from *Uspekhi Mat. Nauk* 18 (1963) no. 5, 13–40).
2. J. Moser, 'On invariant curves of area-preserving mappings of an annulus, *Nachr. Akad. Wiss. Göttingen* 1 (1962) 1–20.
3. I. C. Percival, 'Chaos in hamiltonian systems', *Proc. Roy. Soc. London* A413 (1987) 131–143.
4. H. Poincaré, *Les méthodes nouvelles de la méchanique céleste* (Gauthier-Villars, Paris, 1892).

ENTROPY IN ERGODIC THEORY – THE INITIAL YEARS

William Parry

The years 1954–59 were especially fruitful even for one renowned for so many singular contributions to mathematical research. There was profusion as always (in fact 52 separate items are listed in Kolmogorov's bibliography, including books and other expository work); but more important, these were the years in which he:

(i) proposed and solved his famous perturbation theorem for Hamiltonian systems (the beginning of KAM – Kolmogorov, Arnold, Moser – theory [1954a, b] [**4**, **10**];

(ii) extended Shannon's work to continuous state processes (along with Dobrushin, Gelfand, Pinsker, Yaglom and others) (see for example [**13**]);

(iii) introduced ε-entropy and other related ideas into the problem of measuring the massiveness of function spaces [1959b];

(iv) made the first major inroads to Hilbert's 13th problem concerning the representability of continuous functions of several variables by the superposition of functions of fewer variables (the reduction from three to two variables was made by the 3rd year student Arnold [**5**] after which Kolmogorov reduced the variables from two to one together with addition) [1957b, 1959b];

(v) introduced entropy into ergodic theory and solved the principle outstanding problem of that time [1958f, 1959a].

All but (i) above are concerned with some variation on the theme of entropy.

It is a specifically *ergodic theoretic* point of view that holds together the problems of stochastic processes and those of dynamical systems. But it was a specifically *Kolmogorov* point of view that enabled him to make a smooth transition between these areas and Hilbert's 13th problem in the same period. The problems of carrying forward Shannon's ideas on entropy from finite state processes to continuous state processes are natural enough and indeed had been probed by Shannon and his coworkers [**18**]. But the successful application of these ideas to *formally* deterministic systems represented a significant leap forward.

Or perhaps one should speak of Kolmogorov's *lead* forward. For although Kolmogorov's two papers [1958f, 1959a] prepared the basic groundwork for future developments, it was Sinai who shaped these ideas into a more serviceable form.

Perhaps today when we witness a surfeit of papers and a number of significant results relating to 'chaos' and 'strange attractors' the idea that deterministic dynamical systems can exhibit randomness or stochasticity is no longer surprising, but there is no doubt that present day discoveries in these areas are the florescence of the seed sown by Kolmogorov.

It would be too much to claim that Kolmogorov foresaw the main features of development in dynamical systems over the subsequent 30 years, but his papers on entropy in its various guises, combined with his emerging ideas on complexity, do suggest a totally new point of view from which to understand dynamics. Today we might say that, Laplace notwithstanding, a *human* scale of observation provides even deterministic systems with the main features of stochastic processes and thereby important invariants. The invariants are computed from observations of the system through a finitely partitioned lens, so to speak.

Ergodic theory is concerned with the behaviour of discrete or continuous time dynamical systems for which there is a measure (volume) which remains invariant through the passage of time. In the discrete case such a dynamical system amounts to a single measure preserving transformation $T: X \to X$ together with its iterates $T^n = T^{n-1} \circ T$. The measure m (which we assume to be a *probability* measure, that is, $m(X) = 1$) enjoys the invariance property

$$\int f(x)\, dm = \int f(Tx)\, dm$$

whenever f is integrable.

In the 1930s von Neumann **[23]** had shown that ergodic transformations T with *pure point spectra* (those for which the operator $U_T: L^2(X) \to L^2(X)$, $U_T f = f \circ T$ has pure point spectrum) are completely characterised by their spectra. Subsequently some progress was made in the analysis of transformations with mixed spectra but an understanding of transformations with continuous spectra, and in particular with Lebesgue spectra, was completely lacking. This meant that the all-important transformations from probability theory such as independent (Bernoulli) shifts and Markov shifts remained mysterious from an invariant point of view, for (apart from the period in the Markov case) they are spectrally indistinguishable one from another.

In retrospect it is clear that information theory in the hands of Shannon, McMillan and others, through their approximation and coding procedures, was developing precisely the required machinery for the problem left open by von Neumann. But there were many other directions suggested by these procedures. Suffice it to say that continuous state stochastic processes needed an information-theoretic treatment, which occupied Kolmogorov, Dobrushin, Gelfand, Pinsker and Yaglom **[13]**. Kolmogorov also saw the possibility of an entropy-theoretic analysis of large sets such as various function spaces [1959b], and went on from there to exploit these ideas in a major attack on Hilbert's 13th problem. In short Kolmogorov, in these years, moved from one important problem to another.

In 1958 Kolmogorov published a solution to the main outstanding problem of ergodic theory. If $\alpha = (A_1, A_2, \ldots)$ is a partition of X, the entropy of α is defined as

$$H(\alpha) = -\sum_i m(A_i) \log m(A_i).$$

One can generate finer and finer partitions α^n with the aid of T. The sets in α^n are typically $A_{i_0} \cap T^{-1} A_{i_1} \cap \ldots \cap T^{-n} A_{i_n}$ and one proves that $H(\alpha^n)/n$ converges to a limit $h(T, \alpha)$ (possibly infinity).

Kolmogorov's initial definition of entropy [1958f] concerned only discrete time or continuous time dynamical systems satisfying the 0–1 law, which he called quasi-regular automorphisms and flows, respectively. (They are now known as K-automorphisms and K-flows.) These systems are analogues of regular stationary stochastic processes. He was to stress in his subsequent paper [1959a] that the definition given in [1958f] was principally motivated by his desire to capture entropy for continuous time systems. His first note suffered from a severe but not fatal flaw, and even his second failed to give the most efficient definition of entropy.

Kolmogorov's second, *improved*, definition [1959a] is as follows: if $H(\alpha) < \infty$ and if α is a *generator* (that is, α generates the full σ-algebra under iteration) then the entropy of T is

$$h_1(T) = h(T, \alpha).$$

In the same paper he proves the vital invariance principle that $h_1(T)$ is independent of the generator α. The same result, however, was also established by Sinai **[19]** using *his* definition of entropy, the definition which is now universally preferred:

$$h(T) = \sup h(T, \alpha)$$

where the supremum is taken over all *finite* partitions. (Kolmogorov had earlier obtained this result via the approach of [1959a], but had favoured the method of [1958f], as it seemed to cover continuous and discrete time dynamics simultaneously.) As Rohlin showed, one obtains the same quantity – the Kolmogorov–Sinai invariant

– if one takes the supremum over all countable α with $H(\alpha) < \infty$. Rohlin also showed that the two definitions coincided for (at least) ergodic measure preserving transformations [15].

In any case, the invariance of $h(T)$ (or $h_1(T)$) gives an immediate solution to the problem bequeathed by von Neumann, for it is easy to show that a Bernoulli shift T based on the probability vector $(p_1, \ldots, p_k)$ has entropy

$$h(T) = -\sum_{i=1}^{k} p_i \log p_i,$$

and so there are infinitely many Bernoulli shifts which are not isomorphic although they have the same Lebesgue spectrum of infinite multiplicity.

Thus Kolmogorov solved this hitherto recalcitrant problem. More important, entropy made its entrance into ergodic theory and dynamical systems. Thirty years on, the potency of this idea is far from spent. From the beginning, especially with the work of Sinai [21], Anosov]3], and others, it was clear that entropy was an essential concept for a complete understanding of the classical work of Hedlund, Hopf and Morse on geodesic flows on manifolds of negative curvature. In fact Sinai [20] made an early connection between these geodesic flows and the other important notion referred to above, namely that of quasi-regularity for flows and automorphisms.

As we have indicated, Kolmogorov's *improved* definition of entropy is not the most efficient, and his first is based on a fallacy which invalidated three of the four stated theorems! But the insight and methods were not to be confounded by such a trifle. Kolmogorov had grounded his reasoning (implicitly) on the *false* statement: if $\mathscr{A}_1 \supset \mathscr{A}_2 \supset \ldots$ are σ-algebras such that $\bigcap_{n=1}^{\infty} \mathscr{A}_n = \mathscr{N}$ (the 0–1 σ-algebra) then for any other σ-algebra $\mathscr{B}$, $\bigcap_{n=1}^{\infty} (\mathscr{A}_n \vee \mathscr{B}) = \mathscr{B}$. The mistake was pointed out by Rohlin who used a counterexample based on the 6-adic rationals. (See [1959a].)

Was there a demon-virus spreading this fallacy in 1958? In the same year Wiener, another stochastic giant, made precisely the same error in his lectures 'Non-linear problems in random theory' [25] when attempting to prove that a certain class of continuous state processes can always be represented as a function of an independent process. (This was actually joint work with Kallianpur done in Calcutta in 1956. See [9]. The error in this case was discovered by Rosenblatt [16].)

To return to Kolmogorov's notion of the entropy of a measure preserving transformation, or rather to Sinai's generally accepted definition, the first point to be made is that although entropy can easily be defined for a finite state stationary process by virtue of the fact that the states are *specified*, Kolmogorov's initial paper and its correction (not to mention verbal accounts of Kolmogorov's seminars of that time) give ample evidence that the search for an appropriate definition of entropy of a dynamical system (lacking an intrinsic, invariant, state structure) presented significant problems. These problems are closely related to the natural prejudice one had in viewing stochastic processes as 'random' and dynamical systems as 'deterministic'. There are of course (singular) stochastic processes which warrant their description as deterministic, but what can one mean by the random behaviour of a dynamical system given by an autonomous differential equation or by the iteration of a 'known' transformation?

If we restrict ourselves to a measure preserving transformation T and make *human* observations of the dynamical system generated by iteration, then we effectively partition the space X into regions $(A_1, \ldots, A_k) = \alpha$ and observe the behaviour of a point x as it moves in discrete time. That is, we record which regions the iterations of x move into. This is surely the case as our record of the movement is not likely to be infinitely precise. More often than not we are interested in a specific physical aspect of the system and the function defining this observable can only be approximately represented. It is precisely this constraint on our knowledge that produces for us a random process (defined by α) or that, more recently, has prompted the designation 'chaos' – especially when combined with the effects of 'exponential divergence' of trajectories.

A concrete example can be given of how the erroneous distinction between stochastic processes and differentiable measure-preserving maps worked on the minds of those who were there at the beginning of entropy in ergodic theory. Sinai **[22]** has related his attempt to compute the entropy of an ergodic group automorphism of a two-dimensional torus. He was not alone, at that time, in supposing that the answer must be zero. However, on showing his sketches of the geometry of the system (no doubt the now well known complementary expansions and contractions) to Kolmogorov, he was persuaded that, on the contrary, the map should have positive entropy. Once he had changed his point of view, he proceeded rather quickly to compute the correct positive value $\log|\beta|$ where β is the eigenvalue with maximum modulus. This is an interesting instance of Kolmogorov's penetrating insight and formidable intuition. In fact he even provided Sinai with the most appropriate partition with which to compute the entropy. Other instances occur in [1958f], despite its awesome gaffe; for example, the stress on the importance of K-systems; the warning that processes with independent increments would not provide examples of flows with *finite* entropy; the 'artificial', but 'interesting', examples of flows with an arbitrary positive and finite value which he computes heuristically and which (if I understand correctly) can be represented as a *suspension* over a Bernoulli shift – a type of flow which became important later in connection with hyperbolic flows.

There were many immediate problems to be undertaken following [1958f] and [1959a], the most pressing being the conjectures or implied conjectures. How does Kolmogorov's definition relate to Sinai's modification? This was clarified by Rohlin **[15]**, who proved that if T is aperiodic and $h(T) < \infty$ then there exists a generator α with $H(\alpha) < \infty$. Thus for all important cases $h_1(T) = h(T)$. How do we solve Kolmogorov's problem of defining entropy for a flow $\{T_t : t \in \mathbb{R}\}$? This was achieved by Abramov **[1]**, who showed that $h(T_t) = |t|\,h(T_1)$ so that the natural definition is $h(T_1)$. Kolmogorov asserted that a (quasi-regular) K-automorphism has countable Lebesgue spectrum (in the orthocomplement of the constants). He had proved this (in the form of regular stationary processes) in **[6]** (see also **[14]**) and Sinai proved the analogue for K-flows **[20]**.

There were many remarkable results to be proved in later years. Ergodic theory increasingly converged on its origins – differentiable dynamics and the foundations of statistical mechanics. The key notions of *topological entropy* and *Lyapunov exponents* **[12]** played their parts here. As to the former (initially defined by Adler, Konheim and McAndrew **[2]**), a more serviceable definition was provided by Dinaburg **[8]** and Bowen **[7]** inspired by Kolmogorov's definition of ε-entropy and ε-capacity (but see also Appendix 7 of **[18]**). The closely related notion of *pressure* was then defined and explored by Ruelle **[17]** and Walters **[24]**.

But to provide a reasonable 'conclusion' to this brief account, one cannot do better than to quote Ornstein's definitive complement **[11]** to Kolmogorov's papers: two Bernoulli shifts with the same entropy are isomorphic.

Kolmogorov's introduction of entropy into ergodic theory gave rise to an exponential growth of activity in measure-theoretic and smooth dynamics. Ornstein's theorem gave a further impetus to research in this area. The combined effect was a renewal of the broad direction opened up by Birkhoff and von Neumann in the 1930s, enabling a rich variety of contacts with other areas to be made.

Postscript. Roy Adler informs me that before news of the work of Kolmogorov and Sinai had arrived in the USA, but after it had been done, Kakutani had posed the problem of distinguishing the Bernoulli shifts based on $(\frac{1}{2},\frac{1}{2})$ and $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ using entropy to his student R. Scoville. It seems that *one* of the reasons for the lack of success was that Scoville used only 2-set partitions. Apparently von Neumann also thought that these shifts could be distinguished with the aid of Shannon's ideas. As Kolmogorov has acknowledged, D. Z. Arov had attempted to use entropy in ergodic theory in his Odessa dissertation (1957) but without conclusive results.

I should like to thank Professor Ya. G. Sinai for several helpful discussions. I am also grateful to Ann Dowker who kindly translated one of Kolmogorov's papers for me.

References

1. L. M. ABRAMOV, 'On the entropy of a flow', *Dokl. Akad. Nauk SSSR* 128 (1959) 873–875; *Amer. Math. Soc. Transl.* (2) 49, 167–170.
2. R. L. ADLER, A. G. KONHEIM and M. H. MCANDREW, 'Topological entropy', *Trans. Amer. Math. Soc.* 114 (1965) 309–319.
3. D. V. ANOSOV, 'Geodesic flows on closed Riemannian manifolds of negative curvature', *Tr. Mat. Inst. Akad. Nauk SSSR* 90 (1967).
4. V. I. ARNOLD, 'Proof of a theorem of A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian', *Uspehi Mat. Nauk SSSR* 18 (1963) 5 (113) 13–40: *Russian Math. Surveys* vol. 18 no. 5 (1963) 9–36.
5. V. I. ARNOLD, 'On the representation of continuous functions of three variables by superposition of continuous functions of two variables', *Mat. Sbornik* 48 (90) no. 1 (1959) 3–74.
6. V. I. ARNOLD and L. D. MESHALKIN, 'Kolmogorov's seminar on selected problems of analysis', *Uspekhi Mat. Nauk SSSR* 15 (4) (1960) 247–259.
7. R. BOWEN, 'Entropy for group endomorphisms and homogeneous spaces', *Trans. Amer. Math. Soc.* 153 (1971) 401–414.
8. E. I. DINABURG, 'The relation between topological entropy and metric entropy', *Dokl. Akad. Nauk. SSSR* 190 (1970); *Sov. Math. Dokl.* 11, no. 1 (1970) 13–16.
9. P. MASANI, 'Wiener's contributions to generalised harmonic analysis, prediction theory and filter theory', *Bull. Amer. Math. Soc.* 72 (1966) 73–125.
10. J. MOSER, *Stable and random motions in dynamical systems.* Annals of Math. Studies 77 (Princeton University Press, 1973).
11. D. S. ORNSTEIN, 'Bernoulli shifts with the same entropy are isomorphic', *Adv. in Math.* 4 (1970) 337–352.
12. YA. B. PESIN, 'Characteristic Lyapunov exponents and smooth ergodic theory', *Russian Math. Surveys* 32, no. 4 (1977) 55–114.
13. M. S. PINSKER, *Information and information stability of random variables and processes* (Holden-Day, 1964).
14. V. A. ROHLIN, 'Exact endomorphisms of a Lebesgue space', *Izv. Nauk. SSSR Ser. Mat.* 25 (1961) 499–530; *Amer. Math. Soc. Transl.* (2) 39 (1964) 1–36.
15. V. A. ROHLIN, 'Generators in ergodic theory', *Vest. Leningrad Univ. Mat. Mah. Ast.* (1963) 26–32.
16. M. ROSENBLATT, 'Stationary processes as shifts of functions of independent random variables', *J. Math. Mech.* 8 (1959) 665–681.
17. D. RUELLE, 'Statistical mechanics on a compact set with Z^{ν} actions satisfying expansiveness and specification', *Bull. Amer. Math. Soc.* 78 (1972) 988–991.
18. C. E. SHANNON and W. WEAVER, *The mathematical theory of communication* (University of Illinois Press, Urbana, 1963).
19. YA. G. SINAI, 'On the concept of entropy for dynamical systems', *Dokl. Akad. Nauk. SSSR* 124 (1959) 768–771.
20. YA. G. SINAI, 'Dynamical systems with countably multiple Lebesgue spectrum I', *Izv. Akad. Nauk. SSSR* 25 (1961) 899–924; *Amer. Math. Soc. Transl.* (2) (1964) 83–110.

21. Y. G. Sinai, 'Probabilistic ideas in ergodic theory', *Internat. Congress Math. Stockholm* (1962) 540–559.
22. Ya. G. Sinai, 'About A. N. Kolmogorov's work on the entropy of dynamical systems', *Ergodic Theory Dynamical Systems* 8 (1988) 501–502.
23. J. von Neumann, 'Zur Operatoren methode in der klassischen mechanik', *Ann. of Math.* 33 (1932) 587–642.
24. P. Walters, 'A variational principle for the pressure of continuous transformations', *Amer. J. Math.* 97 (1976) 937–971.
25. N. Wiener, *Non-linear problems in random theory* (M.I.T. Press, 1958).

KOLMOGOROV AND THE COMPLEXITY OF ALGORITHMS

A. A. Razborov

1. *Kolmogorov–Uspenskii machines*

Andrei Nikolaevich (ANK in what follows) appears first to have become interested in the theory of algorithms in 1951 when he suggested to his student V. A. Uspenskii (now professor at the Moscow State University) the study of a new definition of the notion of an algorithm. The main ideas and results of their research were summarised in the talk 'On the notion of algorithm' given by ANK on 17 March 1953 at a session of the Moscow Mathematical Society. (For a summary of this see [1953a] and the joint paper [1958i].) Later such algorithms became known as Kolmogorov–Uspenskii algorithms. Let us sketch their definition. A Kolmogorov–Uspenskii algorithm (or machine) operates stepwise on labelled graphs of a special kind. At each moment there is a distinguished vertex. One working step of a Kolmogorov–Uspenskii machine consists of rebuilding a neighbourhood of the distinguished vertex, following the instructions of the machine. There is an important restriction that the diameter of the neighbourhood is to be bounded from above by a constant depending only on the machine under consideration. The machine halts when the current neighbourhood coincides with one on a list given in advance. The result of the computation is determined by this neighbourhood.

Before ANK's approach, many different methods had been suggested for making precise the intuitive notion of an algorithm, for example, the Herbrand–Gödel calculus, the Turing machines, recursive functions, Markov algorithms, and so on. All are pairwise equivalent, and the famous Church thesis states that *any reasonable definition should be equivalent to those above.* But simulations of algorithms of one kind by those of another kind are not direct: they make use of auxiliary codings. It seems that ANK was the first to realise that one has to be interested in the simplicity and naturalness of such simulations, rather than in the mere fact of their existence. ANK foresaw in this way many of the ideas of modern complexity theory. Here is what he and Uspenskii wrote in [1958i].

We want to emphasise that the question is not the reducibility of an arbitrary algorithm to an algorithm in our sense ⟨the Kolmogorov–Uspenskii algorithms⟩ ... *but that any algorithm essentially fits the definition proposed.*

It was proved in [1958i] that the set of functions computed on a Kolmogorov–Uspenskii machine is exactly the set of recursive functions (this is additional evidence in support of the Church thesis), and, what is more important, strong arguments were given demonstrating that all previously known definitions of the concept 'algorithm' can be simulated by Kolmogorov–Uspenskii machines in a direct 'local' way.

These ideas of ANK were extended by A. Schönhage [5]. He proved, in particular, the possibility of simulating any multi-dimensional Turing machine by a Schönhage machine (this being just a Kolmogorov–Uspenskii machine of a special kind) in real time. This is a reasonable refinement of the intuitive notion of a 'direct' simulation in the sense of Kolmogorov.

2. *The algorithmic theory of information*

It is clear from intuition that the word 111111111111 contains less information than 3.1415926535, while the latter, in turn, contains less information than a 'random word' consisting of twelve digits. But how does one make this intuition precise? ANK was one of the first to realise that it is impossible to do this within the framework of 'pure' probability theory without involving the theory of algorithms. In his first paper [1963d] devoted to this question he wrote:

We can show that in sufficiently large populations the distribution of a property may be such that the frequency of its occurrence will be almost the same for all sufficiently large sub-populations, when the law of choosing these is sufficiently simple. Such a conception in its full development requires the introduction of a measure of the complexity of the algorithm.

The simplest way of understanding the main ideas of [1963d] is to consider an imaginary experiment involving card-guessing. Assume that there is a finite set of cards lying back-to-front on a table, with the successive digits of an unknown binary word written on their hidden faces. Assume also that an algorithm for turning over the cards is given, which at each step (perhaps using information gained from previous steps) points to a card not used before and either guesses the symbol written on that card or passes. The card is then turned over and, if a guess has been made, the player checks whether or not it was correct.

Extending old ideas of von Mises, ANK proposed that one should consider a binary word to be 'random' if any 'sufficiently simple' algorithm when applied to this word does not lead to a proportion of correct answers that is 'essentially more' than 50 per cent. Roughly speaking, ANK showed in [1963d] that, if the total number of algorithms under consideration is not 'too large', then such 'random words' do exist. He then wrote, in the foreword to the Russian translation of [1963d]:

The main differences from the papers by von Mises are the entirely finite nature of the whole framework, and the introduction of a quantitative estimate for the stability of frequencies.

It was the introduction of this quantitative approach that led ANK to a precise definition of a measure of randomness for finite objects.

It seems that the paper [6] published in 1964 by Solomonoff was one of the first publications containing a project for reconstructing information theory. ANK arrived at a similar idea in 1963–64 without having known about Solomonoff's research, and he published his first paper [1965a] on this topic in the beginning of 1965. In that paper he investigated the concept of the complexity of a finite object y with respect to a given finite object x. Afterwards this concept was called the *Kolmogorov complexity.*

Given a partially recursive function $\phi(p,x)$ (throughout the following p, x, and y are binary words), the Kolmogorov complexity $K_\phi(y|x)$ is defined by

$$K_\phi(y|x) = \min_{\phi(p,x)=y} |p| \quad \text{if such words } p \text{ exist,}$$

$$K_\phi(y|x) = \infty \quad \text{otherwise.}$$

Informally $K_\phi(y|x)$ is the minimal possible length $|p|$ of a program p that yields y given x.

The main theorem proved in [1965a] states that there exists a partially recursive function $A(p, x)$ such that for every other partially recursive function $\phi(p, x)$, we have

$$K_A(y|x) \leqslant K_\phi(y|x) + C_\phi,$$

with a constant C_ϕ that does not depend on x and y. For any A and B both satisfying the conditions of the theorem we have

$$|K_A(y|x) - K_B(y|x)| \leqslant C_{A,B}$$

and so, up to an additive constant, we can speak about the Kolmogorov complexity $K(y|x)$ irrespective of the choice of A.

The most important particular case is that of an empty x, and then $K(y) = K(y|x)$ is called the *(absolute) Kolmogorov complexity* of y.

It is impossible to overestimate the importance of these ideas and results. Here we sketch only a few of the later developments.

ANK involved his pupils in the elaboration of his ideas, especially L. A. Levin and P. Martin-Löf. The Kolmogorov complexity $K(y|x)$ can be regarded as the entropy of y given x. In accordance with this it is sometimes denoted by the letter H instead of K. Extending the analogy one defines the *amount of information* $I(y:x)$ contained in y relative to x by the formula

$$I(y:x) = H(x) - H(x|y).$$

Kolmogorov and Levin proved in [1969a] that two fundamental formulae of information theory can be retained in the new algorithmic framework in the following approximate form:

$$|I(y:x) - I(x:y)| = O(\log_2 H(x, y)),$$

$$H(x, y) = H(x) + H(y|x) + O(\log_2 H(x, y)).$$

Starting from the ideas of Kolmogorov, Martin-Löf [3] proposed in 1966 a definition of a random (infinite) binary sequence. ANK conjectured in [1969a] that there should be an alternative definition of this notion in terms of the Kolmogorov complexity of the finite left-hand segments of the given sequence, and pointed out some difficulties that arise in pursuing this course. These difficulties were overcome independently by Levin [**2**] and C. P. Schnorr [**4**]. It turned out that a sequence $x_1, x_2, \ldots, x_n, \ldots$ is random in Martin-Löf's sense if and only if

$$KM((x_1, x_2, \ldots, x_n)) = n + O(1).$$

Here KM is a slight modification of the Kolmogorov complexity – the so-called monotone Kolmogorov complexity.

The reader wishing to see further details is referred to [**1**] and to [1983d]. The algorithmic theory of information is currently being actively studied by several young pupils of Andrei Nikolaevich: E. A. Asarin, A. Kh. Shen', V. G. Vovk, and others.

References

1. A. K. Zvonkin and L. A. Levin, 'The complexity of finite objects, and the basing of the concepts of information and randomness on the theory of algorithms', *Russian Math. Surveys* 25 (1970) no. 6, 83–124.
2. L. A. Levin, 'The concept of a random sequence', *Dokl. Akad. Nauk SSSR* 212 (1973) no. 3, 548–550.

3. P. MARTIN-LÖF, 'The definition of random sequences', *Inform. and Control* 9 (1966) 602–619.
4. C. P. SCHNORR, 'Process complexity and effective random tests', *J. Comp. and Systems Science* 7 (1973) 376–388.
5. A. SCHÖNHAGE, 'Storage modification machines', *SIAM J. Comput.* 9 (1980) 490–508.
6. R. SOLOMONOFF, 'A formal theory of inductive inference', *Inform. and Control* 7 (1964) 1–22.

THE WORK OF KOLMOGOROV ON COHOMOLOGY

C. A. ROBINSON

In the mid-1930s Kolmogorov played a substantial part in the early development of cohomology theory. It happened that almost exactly the same discoveries were made independently and simultaneously by J. W. Alexander in the USA, and were published by him in English. As a result, Kolmogorov's work on this subject has been less well known in English-speaking countries than it deserves to be. Both Kolmogorov and Alexander lectured on their discoveries at the International Topology Conference at Moscow in 1935.

The first aspect of Kolmogorov's work was begun in 1934 and concerns duality theory in cell complexes [1936f]. At this time, the topological invariance of homology groups was a standard result, but cohomology existed only as the theory of 'pseudo-cycles' in a section of Lefschetz's 1930 text on topology. For any finite cell complex K, Kolmogorov in his paper introduces the chains, cochains, homology and cohomology groups of K, with coefficients in a locally compact abelian group. (Homology and cohomology groups are respectively called 'Betti u-groups' and 'Betti o-groups' in his work.) He shows that $H^r(K;A)$ and $H_r(K;B)$ are dual groups if A and B are Pontryagin duals of each other. When L is a subcomplex of some cellular decomposition of an n-sphere, he proves the celebrated duality theorem of Alexander and Kolmogorov:

$$\tilde{H}_{r-1}(L) \approx H^{n-r}(S^n - L), \quad 1 < r < n.$$

Kolmogorov's next contribution was the definition of cohomology groups for arbitrary locally compact topological spaces. He published this in four notes to the *Comptes Rendus* [1936 l, m, n, o]. The same discovery (for compact spaces) was made independently by Alexander, and the result is known as Alexander–Kolmogorov cohomology. One can think of it as dual (in Pontryagin's sense) to Vietoris homology, which is obtained by considering a finite set of nearby points as a 'simplex' in a space.

The construction of the cohomology groups can be outlined as follows. If X is a locally compact space, Kolmogorov considers the alternating functions from X^{r+1} into an abelian coefficient group $\mathscr{J}$. The group $F^r = F^r(X;\mathscr{J})$ of r-cochains on X is obtained from these by identifying two functions if they agree on some neighbourhood of the diagonal in X^{r+1}. The coboundary operator $g_0: F^{r-1} \to F^r$ is defined by

$$(g_0 f)(p_0, p_1, \dots, p_r) = \sum_{i=0}^{r} (-1)^i f(p_0, \dots, p_{i-1}, p_{i+1}, \dots, p_r).$$

Since $g_0 \circ g_0 = 0$, one can define the cohomology group $H^r(X;\mathscr{J})$ as the subquotient $(\ker g_0)/(\operatorname{im} g_0)$ of F^r. Kolmogorov relates this to the corresponding homology theory, which generalizes that of Vietoris. He introduces relative homology and cohomology groups, develops some of their standard properties, and then deduces a variety of forms of the duality theorem for the homology and cohomology of a closed subset and its complement in a manifold.

The first major advantage of cohomology over homology is the existence of the ring structure. In the case of a manifold M, the product in $H^*(M)$ corresponds under Poincaré duality to the geometrical operation of intersecting cycles. In [1936h], Kolmogorov introduces a precursor of the product we use today. His construction relates to the rational cohomology of an arbitrary simplicial complex. In order to define the modern product, one introduces an ordering of the vertices. The complexes in this paper are unordered, and so Kolmogorov obtains a ring structure in which the product of an r-dimensional class with an s-dimensional one is larger than the modern one by a factor of $(r+s)!/r!s!$. This important work on the cohomology ring was also announced in Kolmogorov's 1935 Moscow lecture.

I should like to thank Professor G. W. Whitehead for some helpful comments on Kolmogorov's work in this field.

KOLMOGOROV'S CONTRIBUTIONS TO THE THEORY OF STATIONARY PROCESSES

PETER WHITTLE

An interest in stationary processes is incidental to a considerable part of Kolmogorov's work, but his direct and classic contributions to the subject were made over a relatively brief span of time. The note published in *Comptes Rendus* [1939e] indicated that these matters were stirring in his mind. Two short papers [1940b, c] published the following year revealed that he had registered the connection between stationary processes and Hilbert space. In September/October of that year he wrote his classic study [1941a] on stationary processes in Hilbert space, and then continued in November with a more explicit account [1941b] of the interpolation and extrapolation of such processes.

His interest was evidently stirred by Wold's work [7], to which he made repeated reference. The principal result of this work was the formula which immediately became known as Wold's decomposition. This expresses the variable $x(t)$ of a scalar-valued stationary process as the sum of a *singular* component (linearly predictable from its remote past) and a one-sided moving average of white-noise variables (later to be known as the *innovations* of the process).

In his *Comptes Rendus* note Kolmogorov simply states his evaluations of the variance of interpolation error and k-step extrapolation (prediction) error. These evaluations are both celebrated and satisfying: the interpolation variance and the one-step prediction variance are respectively the harmonic and geometric means of the spectral density function. The case of zero prediction error was Wold's singular case. It is now plain to us from this note that Kolmogorov had essentially solved the prediction problem for stationary processes; the note was perhaps too bare and cryptic for this to be evident at the time.

The principal results established in the literature when Kolmogorov began his major advance in 1940 were those by Khinchin [3] and Cramér [1] on the spectral representation of the autocovariance function, and Wold's decomposition. Kolmogorov's insight was to see a scalar random variable of finite second moment as an element of a Hilbert space H, and the expected product of two such random variables as defining the inner product of the two elements. Suppose that all the members of a stationary sequence $\{x(t)\}$ are elements of H, and, indeed, that each element of H can be seen as a member of one of a number of jointly stationary sequences. The operator U inducing the time translation $Ux(t) = x(t+1)$ then leaves the inner product invariant. That is, U is recognised as a unitary operator on H, and a substantial known theory can be brought to bear. In particular, the spectral representation of U is essentially a Fourier representation, and this implies corresponding Fourier representations of the autocovariance and of the random function $x(\cdot)$ itself.

The two 1940 papers are in fact concerned with processes whose increments are stationary. However, the results deduced do include as a special case the spectral representation of the stationary random function $x(\cdot)$ (as the Fourier–Stieltjes transform of a process of uncorrelated increments). While this result was implicit in earlier abstract work by von Neumann [4] and Wintner [6], this was the first realisation of its implications in the stationary process context, certainly preceding the oft-quoted work of Cramér [2]. An even longer sighting into the future on Kolmogorov's part is the beginning of a study of self-similar processes, with its recognition of the special role of power-law autocovariance functions.

Kolmogorov's major work in this area [1941a] gave a complete analysis of stationary processes using the Hilbert space formulation, and integrated Wold's results into an analytic framework associated with the spectral representations. He first introduced the idea of *subordination*: that one stationary process is subordinate to another if it can be represented as a moving average of that other. He translated the idea of subordination into spectral terms; also the decomposition of a process into mutually orthogonal components. Proceeding in this way, he deduced the Wold decomposition, with an interpretation of the components in spectral terms.

The article ends with a treatment of linear least-square interpolation and extrapolation, deriving the results asserted in *Comptes Rendus*, and giving a complete spectral characterisation of what we would now term the purely deterministic and purely non-deterministic cases. This treatment is given greater body and detail in the immediately following paper [1941b].

Kolmogorov and Wiener [5] are generally given joint credit for the development of the prediction theory of stationary processes. This surely constitutes insufficient recognition of Kolmogorov's clear ten-year priority. On the other hand, one can commiserate with Wiener in his misfortune. The war that gave him the stimulus to work on these problems both delayed his own publication (on security grounds) and left him unaware of Kolmogorov's results. The contrast in the approaches of the two is interesting. Wiener quickly cast the problem into analytic form; he derived conditions for the optimal predictor and converted these into a Wiener–Hopf problem. Kolmogorov took a more directly temporal and statistical approach; just the simple Gram-Schmidt orthogonalisation which Wold had employed and which led to the fruitful concept of an innovation. Of course, the two approaches had to converge at some point. Kolmogorov saw that the generating function of the coefficients in the innovations representation constituted a canonical factor of the spectral density function, that it was the limit on the unit circle of a function analytic within, and that it had what we should now term the minimum-phase property. When one compares this work with all that has been done since one is impressed by its ease and economy. On the other hand, it is just this economy which has made the work a 'mathematician's' work; Wiener's writings are better known by practitioners, not only because they are located in the Western literature and because they make specific reference to applications, but also because the labour of analysis is evident, and so the achievement itself more evident.

One might say that some concession is made in this direction in an article which Kolmogorov published a few years later [1947f] in a commemorative volume. This article gave considerable attention to the spectral representation of the random function itself, which had played a surprisingly slight role in the 1941 memoir. It also made clear the practical significance of these concepts by, for example, deducing the spectral density of a process generated by driving a differential equation with white noise.

References

1. H. Cramér, 'On the theory of stationary random processes', *Ann. of Math.* 41 (1940) 215–230.
2. H. Cramér, 'On harmonic analysis in certain functional spaces', *Ark. Mat. Astr. Fys.* 28B (1942).
3. A. Khinchin, 'Korrelationstheorie der stationären stochastischen Prozesse', *Math. Ann.* 109 (1934) 604–615.
4. J. von Neumann, 'Allgemeine Eigenwerttheorie Hermitischer Funktionaloperatoren', *Math. Ann.* 102 (1929) 49–131.
5. N. Wiener, *Extrapolation, interpolation and smoothing of stationary time series* (Wiley, New York, 1949).
6. A. Wintner, 'Zur Theorie der beschränkten Bilinearformen', *Math. Z.* 30 (1929) 228–282.
7. H. Wold, *A study in the analysis of stationary time series* (Almquist and Wicksell, Uppsala, 1938).

BIBLIOGRAPHY

Complete list of the main works of A. N. Kolmogorov

1923

(a) 'Une série de Fourier–Lebesgue divergente presque partout', *Fundamenta Math.* 4, 324–328.

(b) 'Sur l'ordre de grandeur des coéfficients de la série de Fourier–Lebesgue', *Bull. Acad. Pol. Sci. Ser. A*, 83–86.

1924

(a) 'Une contribution à l'étude de la convergence des séries de Fourier', *Fundamenta Math.* 5, 96–97.

(b) (With G. A. Seliverstov) 'Sur la convergence des séries de Fourier', *C.R. Acad. Sci. Paris* 178, 303–306.

1925

(a) 'La définition axiomatique de l'intégrale', *C.R. Acad. Sci. Paris* 180, 110–111.

(b) 'Sur les bornes de la généralisation de l'intégrale', *Collected works of A. N. Kolmogorov* (Nauka, Moscow, 1985).

(c) 'Sur la possibilité de la définition générale de la dérivée, de l'intégrale et de la sommation des séries divergentes', *C.R. Acad. Sci. Paris* 180, 362–364.

(d) 'Sur les fonctions harmoniques conjuguées et les séries de Fourier', *Fundamenta Math.* 7, 24–29.

(e) 'On the *tertium non datur* principle', *Mat. Sb.* 32, 646–667. (An English translation of this paper can be found in J. van Heijenoort, *From Frege to Gödel: a source book in mathematical logic 1879–1931*, Harvard, 1967, reprinted 1970.)

(f) (With A. Ya. Khinchin) 'Ueber Konvergenz von Reihen, deren Glieder durch den Zufalle bestimmt werden', *Mat. Sb.* 32, 668–677.

(g) (With G. A. Seliverstov) 'Sur la convergence des séries de Fourier', *Atti Accad. Naz. Lincei Rend.* 3, 307–310.

(h) 'Sur une série de Fourier–Lebesgue divergente partout', *C.R. Acad. Sci. Paris* 183, 1327–1329.

1927

(a) 'Sur la loi des grands nombres', *C.R. Acad. Sci. Paris* 185, 917–919.

(b) (With D. E. Men'shov) 'Sur la convergence des séries de fonctions orthogonales', *Math. Z.* 26, 432–441.

1928

(a) 'Operations on sets', *Mat. Sb.* 35, 414–422.
(b) 'Sur une formule limite de M. A. Khintchine', *C.R. Acad. Sci. Paris* 186, 824–825.
(c) 'Sur un procédé d'intégration de M. Denjoy', *Fundamenta Math.* 11, 27–28.
(d) 'Ueber die Summen durch den Zufall bestimmter unabhängiger Grössen', *Math. Ann.* 99, 309–319.

1929

(a) 'Bemerkungen zu meiner Arbeit "Ueber die Summen zufälliger Grössen"', *Math. Ann.* 102, 484–488.
(b) 'General measure theory and the calculus of probabilities', *Proc. Communist Acad., Math. Section* 1, 8–21.
(c) 'Modern arguments about the nature of mathematics', *Nauch. Slovo* 6, 41–54.
(d) 'Ueber das Gesetz des iterierten Logarithmus', *Math. Ann.* 101, 126–135.
(e) 'Sur la loi des grands nombres', *Atti Accad. Naz. Lincei Rend.* 9, 470–474.

1930

(a) 'Sur la loi forte des grands nombres', *C.R. Acad. Sci. Paris* 191, 910–912.
(b) 'Zur topologisch-gruppentheoretischen Begründung der Geometrie', *Nachr. Ges. Wiss. Göttingen Fachgruppe I* (*Mathematik*) 8, 208–210.
(c) 'Untersuchungen über den Integralbegriff', *Math. Ann.* 103, 654–696.
(d) 'Sur la notion de la moyenne', *Atti Accad. Naz. Lincei Rend.* 12, 388–391.

1931

(a) 'Ueber die analytischen Methoden in der Wahrscheinlichkeitsrechnung', *Math. Ann.* 104, 415–458.
(b) 'Sur le problème d'attente', *Mat. Sb.* 38, 101–106.
(c) 'The median method in the theory of errors', *Mat. Sb.* 38, 47–50.
(d) 'Eine Verallgemeinerung des Laplace-Liapounoffschen Satzes', *Izv. Akad. Nauk Otdel. Mat. Estest. Nauk*, 959–962.
(e) 'Ueber Kompaktheit der Funktionenmengen bei der Konvergenz im Mittel', *Nachr. Ges. Wiss. Göttingen* 9, 60–63.

1932

(a) 'Theory of functions of a real variable', *Mathematics in the USSR in the last 15 years* (GTTI, Moscow–Leningrad), pp. 37–48.
(b) 'Sulla forma generale di un processo stocastico omogeneo (un problema di Bruno de Finetti)', *Atti Accad. Naz. Lincei Rend.* 15, 805–808.
(c) 'Ancora sulla forma generale de un processo stocastico omogeneo', *Atti Accad. Naz. Lincei Rend.* 15, 866–869.
(d) 'Zur Deutung der intuitionistischen Logik', *Math. Z.* 35, 58–65.
(e) 'Zur Begründung der projektiven Geometrie', *Ann. of Math.* (2) 33, 175–176.
(f) (With P. S. Aleksandrov) *Introduction to the theory of functions of a real variable* (GTTI, Moscow–Leningrad).

1933

(a) (With P. S. Aleksandrov) *Introduction to the theory of functions of a real variable* (2nd ed., GTTI, Moscow–Leningrad).
(b) *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin).
(c) 'Beiträge zur Masstheorie', *Math. Ann.* 107, 351–366.
(d) (With M. A. Leontovich) 'Zur Berechnung der mittleren Brownschen Fläche', *Phys. Z. Sow.* 4, 1–13.
(e) 'Sulla determinazione empirica di una legge di distribuzione', *Giorn. Ist. Ital. Attuari* 4, 83–91.
(f) 'Ueber die Grenzwertsätze der Wahrscheinlichkeitsrechnung', *Izv. Akad. Nauk SSSR Otdel. Mat. Estest. Nauk*, 363–372.
(g) 'Zur Theorie der stetigen zufälligen Prozesse', *Math. Ann.* 108, 149–160.
(h) 'Sur la détermination empirique d'une loi de distribution', *Uch. Zap. Moskov. Gos. Univ.* 1, 9–10.
(i) 'On the question of the suitability of prediction formulae found by a statistical method', *Zh. Geofiz.* 3, 78–82.

1934

(a) (With I. Ya. Verchenko) 'On points of discontinuity of functions of two variables', *Dokl. Akad. Nauk SSSR* 1, 105–106.
(b) 'Zur Normierbarkeit eines allgemeinen topologischen linearen Raumes', *Studia Math.* 5, 29–33.
(c) (With I. Ya. Verchenko) 'Continuation of research on points of discontinuity of functions of two variables', *Dokl. Akad. Nauk SSSR* 4, 361–362.
(d) 'On the convergence of series in orthogonal polynomials', *Dokl. Akad. Nauk SSSR* 1, 291–294.
(e) 'Quelques remarques sur l'approximation des fonctions continues', *Mat. Sb.* 42, 99–103.

(f) 'On some modern trends in probability theory', *Bull. 2nd All-Union Math. Congress* (*Leningrad, 24–30 June 1934*) (Izdat. Akad. Nauk SSSR), p. 8.
(g) 'Modern mathematics', *Front Nauki i Tekhniki*, no. 5/6, 25–28.
(h) 'Institute of Mathematics and Mechanics of Moscow State University', *Front Nauki i Tekhniki*, no. 5/6, 75–78.
(i) '2nd All-Union Math. Congress', *Sotsialisticheskaya Rekonstruktsiya i Nauka*, no. 7.
(j) 'Zufällige Bewegungen (Zur Theorie der Brownschen Bewegung)', *Ann. of Math.* (2) 35, 116–117.

1935
(a) 'On some modern trends in probability theory', *Proc. 2nd All-Union Math. Congress (Leningrad, 24–30 June 1934)* (Izdat. Akad. Nauk SSSR, vol. 1), pp. 349–358.
(b) 'Deviations from Hardy's formulae in partial isolation', *Dokl. Akad. Nauk SSSR* 3, 129–132.
(c) 'La transformation de Laplace dans les espaces linéaires', *C.R. Acad. Sci. Paris* 200, 1717–1718.
(d) 'Zur Grössenordnung des Restgliedes Fourierschen Reihen differenzierbarer Funktionen', *Ann. of Math.* (2) 36, 521–526.

1936
(a) 'Ueber die beste Annäherung von Funktionen einer gegebenen Funktionenklasse', *Ann. of Math.* (2) 37, 107–110.
(b) *The basic concepts of probability theory* (ONTI, Moscow–Leningrad; Russian translation of [1933b]).
(c) 'Equation', *BSE* (*Great Soviet Encyclopaedia*), Vol. 56, 163–165.
(d) 'Modern mathematics', *Collection of articles on the philosophy of mathematics* (ONTI, Moscow), pp. 7–13.
(e) 'Theory and practice in mathematics', *Front Nauki i Tekhniki*, no. 5, 39–42.
(f) 'Ueber die Dualität im Aufbau der kombinatorischen Topologie', *Mat. Sb.* 1, 97–102.
(g) 'Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zustanden', *Mat. Sb.* 1, 607–610.
(h) 'Homologiering des Komplexes und des lokalbikompakten Raumes', *Mat. Sb.* 1, 701–706.
(i) 'Zur Theorie der Markoffschen Ketten', *Math. Ann.* 112, 155–160.
(j) 'On the Plessner condition for the law of large numbers', *Mat. Sb.* 1, 847–849.
(k) (With P. S. Aleksandrov) 'Endliche Ueberdeckungen topologischer Räume', *Fundamenta Math.* 26, 267–271.
(l) 'Les groupes de Betti des espaces localement bicompacts', *C.R. Acad. Sci. Paris* 202, 1144–1147.
(m) 'Propriétés des groupes de Betti des espaces localement bicompacts', *C.R. Acad. Sci. Paris* 202, 1325–1327.
(n) 'Les groupes de Betti des espaces métriques', *C.R. Acad. Sci. Paris* 202, 1558–1560.
(o) 'Cycles relatifs. Théorème de dualité de M. Alexander', *C.R. Acad. Sci. Paris* 202, 1641–1643.
(p) Preface to the Russian translation of: A. Heyting, *A survey of research on the foundations of mathematics* (ONTI, Moscow–Leningrad), pp. 3–4.
(q) 'Sulla teoria di Volterra della lotta per l'esistenza', *Giorn. Ist. Ital. Attuari* 7, 74–80.

1937
(a) 'Ueber offene Abbildungen', *Ann. of Math.* (2) 38, 36–38.
(b) 'Skew-symmetric quantities and topological invariants', *Proc. Sem. on Vector and Tensor Analysis and their Applications to Geometry, Mechanics and Physics* (GONTI, Moscow–Leningrad), no. 46, pp. 342–347.
(c) 'Markov chains with countably many possible states', *Byul. Moskov. Univ., Mat. i Mekh.* 1, no. 3, 1–16.
(d) (With I. G. Petrovskii and N. S. Piskunov) 'Investigation of the diffusion equation associated with an increase in the quantity of matter and its application to a biological problem', *Byul. Moskov. Univ., Mat. i Mekh.* 1, no. 6, 1–26.
(e) 'On the statistical theory of the crystallization of metals', *Izv. Akad. Nauk SSSR Ser. Mat.* no. 3, 367–368.
(f) 'Ein vereinfachter Beweis des Birkhoff–Khintchineschen Ergodensatzes', *Mat. Sb.* 2, 367–368.
(g) 'Zur Umkehrbarkeit der statistichen Naturgesetze', *Math. Ann.* 113, 766–772.
(h) (With P. S. Aleksandrov) Editorial and Appendix to the Russian translation of: F. Hausdorff, *Mengenlehre* (GONTI, Moscow–Leningrad).
(i) 'The continuum', *BSE*, Vol. 34, 139–140.
(j) 'Grandeurs gauches et invariants topologiques', *Trudy Sem. Vektor. Tenzor. Anal.* 4, 342–347.

1938
(a) (With P. S. Aleksandrov) *Introduction to the theory of functions of a real variable* (3rd revised ed., GONTI, Moscow–Leningrad).
(b) 'Andrei Andreevich Markov', *BSE*, Vol. 38, 152–153.
(c) 'Mathematics', *BSE*, Vol. 38, 359–402.

(d) 'Mathematical induction', *BSE*, Vol. 38, 405–406.
(e) 'Measure', *BSE*, Vol. 38, 831–832.
(f) 'Multidimensional space', *BSE*, Vol. 39, 577–578.
(g) 'Probability theory and its applications', *Mathematics and natural science in the USSR* (GONTI, Moscow–Leningrad), pp. 51–61.
(h) 'On the information section in the first issue of "Uspekhi Matematicheskikh Nauk"', *Uspekhi Mat. Nauk* no. 4, 326–327.
(i) 'A remark concerning the foundations of geometry. (On the question of a new translation of Hilbert's *Grundlagen der Geometrie*)', *Uspekhi Mat. Nauk* no. 4, 347–348.
(j) 'From the Editor (A cycle of articles on the theory of stochastic processes)', *Uspekhi Mat. Nauk* no. 5, 3–4.
(k) 'On analytic methods in probability theory', *Uspekhi Mat. Nauk* no. 5, 5–41 (Russian translation of [1931a]).
(l) 'A simplified proof of the ergodic theorem of Birkhoff and Khinchin', *Uspekhi Mat. Nauk* no. 5, 52–56 (Russian translation of [1937f]).
(m) (With G. M. Fikhtengol'ts and I. M. Gel'fand) 'Some problems in the theory of functions of a real variable', *Uspekhi Mat. Nauk* no. 5, 232–234.
(n) 'On the solution of a biological problem', *Izv. Nauchn. Issled. Inst. Mat. Mekh., Tomsk Univ.* 2, no. 1, 7–12.
(o) 'Une généralisation de l'inégalité de M. J. Hadamard entre les bornes supérieures des dérivées successives d'une fonction', *C.R. Acad. Sci. Paris* 207, 764–765.
(p) Editorial and Preface to the Russian translation of: H. Lebesgue, *Intégrale, longeur, aire* (Ann. Mat. Pura Appl. 7 (1902)) (Uchpedgiz, Moscow.).

1939
(a) (With P. S. Aleksandrov) *Algebra*, Part 1 (Uchpedgiz, Moscow).
(b) 'Orientation', *BSE*, Vol. 43, 342–344.
(c) 'On inequalities between upper bounds of successive derivatives of an arbitrary function on an infinite interval', *Uch. Zap. Moskov. Gos. Univ.* 30 (Matematika), Book 3, pp. 3–16.
(d) (With I. M. Gel'fand) 'Rings of continuous functions on topological spaces', *Dokl. Akad. Nauk SSSR* 22, 11–15.
(e) 'Sur l'interpolation et extrapolation des suites stationnaires', *C.R. Acad. Sci. Paris* 208, 2043–2045.

1940
(a) 'Surface', *BSE*, Vol. 45, 746–748.
(b) 'Curves in a Hilbert space that are invariant under a one-parameter group of motions', *Dokl. Akad. Nauk SSSR* 26, 6–9.
(c) 'The Wiener spiral and some other interesting curves in a Hilbert space', *Dokl. Akad. Nauk SSSR* 26, 115–118.
(d) 'Sergei Natanovich Bernstein (on his sixtieth birthday)', *Izv. Akad. Nauk SSSR Ser. Mat.* 4, 249–260.
(e) 'Valerii Ivanovich Glivenko (1897–1940)', *Uspekhi Mat. Nauk* no. 8, 379–383.
(f) Editorial to the book: V. I. Romanovskii, *Mathematical statistics*, *Uspekhi Mat. Nauk* no. 7, 327–329.
(g) 'On new evidence supporting Mendel's laws', *Dokl. Akad. Nauk SSSR* 27, 38–42.

1941
(a) 'Stationary sequences in a Hilbert space', *Byul. Moskov. Univ. Mat.* 2, no. 6.
(b) 'Interpolation and extrapolation of stationary random sequences', *Izv. Akad. Nauk SSSR Ser. Mat.* 5, 3–14.
(c) 'Points of local topologicity of countably-multiple open maps of compacta', *Dokl. Akad. Nauk SSSR* 30, 477–479.
(d) 'The local structure of turbulence in an incompressible fluid with very large Reynolds numbers', *Dokl. Akad. Nauk SSSR* 30, 301–305.
(e) 'On the logarithmically normal law of distribution of the size of particles under pulverization', *Dokl. Akad. Nauk SSSR* 31, 99–101.
(f) 'On the degeneration of isotropic turbulence in an incompressible viscous fluid', *Dokl. Akad. Nauk SSSR* 31, 538–540.
(g) 'Dissipation of energy under locally isotropic turbulence', *Dokl. Akad. Nauk SSSR* 32, 16–18.
(h) (With P. S. Aleksandrov) 'Properties of inequalities and the concept of approximate calculations', *Mat. v Shkole* no. 2, 1–12.
(i) (With P. S. Aleksandrov) 'Irrational numbers', *Mat. v Shkole* no. 3, 1–15.
(j) 'Confidence limits for an unknown distribution function' (in English), *Ann. Math. Statistics* 12, 461–463.

1942
(a) 'The definition of the centre of dispersion and degree of accuracy for a bounded number of observations', *Izv. Akad. Nauk SSSR Ser. Mat.* 6, 3–32.
(b) 'The equations of a turbulent motion of an incompressible viscous fluid', *Izv. Akad. Nauk SSSR Ser. Fiz.* 6, 56–58. See also *J. Phys.* 6 (1942), 227.

1943
(a) (With P. S. Aleksandrov) *Nikolai Ivanovich Lobachevskii (1793–1843)* (Gostekhizdat, Moscow–Leningrad).
(b) 'Lobachevskii and mathematical thinking in the 19th century', *Nikolai Ivanovich Lobachevskii* (Moscow–Leningrad), pp. 87–100.

1945
(a) 'The number of hits in several shots and the general principles of estimating the effectiveness of a firing system', *Trudy Mat. Inst. Akad. Nauk SSSR* 12, 7–25.
(b) 'Artificial dispersion in the case of defeat by a single hit and dispersions in one variable', *Trudy Mat. Inst. Akad. Nauk SSSR* 12, 26–45.

1946
(a) 'On the question of the law of resistance in turbulent flow in smooth tubes', *Dokl. Akad. Nauk SSSR* 52, 669–671.
(b) 'On the justification of the method of least squares', *Uspekhi Mat. Nauk* 1, no. 1, 57–70.
(c) 'On the justification of the theory of real numbers', *Uspekhi Mat. Nauk* 1, no. 1, 217–219.
(d) 'Newton and modern mathematical thinking', *Moscow University – in memoriam Isaac Newton, 1643–1943* (Izdat. Moskov. Gos. Univ., Moscow), pp. 24–43.
(e) 'Discussion on the article by M. A. Velikanov "Transfer of suspended alluvium by a turbulent flow"', *Izv. Akad. Nauk SSSR* no. 5, 781–784.

1947
(a) 'The development of mathematics in the USSR', *BSE*, Vol. "SSSR", 1318–1323.
(b) 'Mean values', *BSE*, Vol. 52, 508–509.
(c) (With A. A. Petrov and Yu. M. Smirnov) 'A formula of Gauss in the theory of the least squares method', *Izv. Akad. Nauk SSSR Ser. Mat.* 11, 561–566.
(d) (With N. A. Dmitriev) 'Branching stochastic processes', *Dokl. Akad. Nauk SSSR* 56, 7–10.
(e) (With B. A. Sevast'yanov) 'The calculation of final probabilities for branching stochastic processes', *Dokl. Adad. Nauk SSSR* 56, 783–786.
(f) 'The statistical theory of oscillations with a continuous spectrum', *Jubilee collection dedicated to the thirtieth anniversary of the Great October socialist revolution, Izdat. Akad. Nauk SSSR*, vol. 1, 242–252.
(g) 'The role of Russian science in the development of probability theory', *The role of Russian science in the development of world science and culture*, Izdat. Moskov. Gos. Univ., Vol. 1, Book 1, 53–64.

1948
(a) 'The statistical theory of oscillations with a continuous spectrum', *General meeting of the Academy of Sciences of the USSR devoted to the thirtieth anniversary of the Great October socialist revolution* (Izdat. Akad. Nauk SSSR, Moscow–Leningrad), pp. 465–472.
(b) (With B. V. Gnedenko) 'Probability theory', *Mathematics in the USSR, 1917–1947* (Gostekhizdat, Moscow–Leningrad), pp. 701–727.
(c) 'Two theorems on probability: comments', *P. L. Chebyshev, Complete works*, Vol. 3, *Mathematical analysis* (Gostekhizdat, Moscow–Leningrad), pp. 404–409.
(d) 'The structure of complete metric Boolean algebras', *Uspekhi Mat. Nauk* 3, no. 1, 212.
(e) 'A remark concerning Chebyshev polynomials that deviate least from a given function', *Uspekhi Mat. Nauk* 3, no. 1, 216–221.
(f) 'Evgenii Evgen'evich Slutskii (obituary)', *Uspekhi Mat. Nauk* 3, no. 4, 143–151.
(g) Editorial and preface to the Russian translation of: H. Cramér, *Mathematical methods of statistics* (Princeton University Press, Princeton, NJ, 1946), (Izdat. Inostr. Lit., Moscow).
(h) 'Algèbres de Boole métriques complètes', 6th Polish Math. Congress, Warsaw, pp. 22–30.

1949
(a) (With B. V. Gnedenko) *Limit distributions for sums of independent random variables* (Gostekhizdat, Moscow–Leningrad).
(b) 'On the problem of "geometrical selection" of crystals', *Dokl. Akad. Nauk SSSR* 65, 681–684.

(c) 'The solution of a problem in probability theory connected with the question of the mechanism of stratification', *Dokl. Akad. Nauk SSSR* 65, 793–796. (English translation: AMS Transl. no. 53 (1951)).

(d) (With Yu. V. Prokhorov) 'On the sums of a random number of terms', *Uspekhi Mat. Nauk* 4, no. 4, 168–172.

(e) 'The local limit theorem for classical Markov chains', *Izv. Akad. Nauk SSSR Ser. Mat.* 13, 281–300.

(f) 'Fundamental problems of theoretical statistics', *Proc. 2nd All-Union Conference on Mathematical Statistics (Tashkent, 27 Sept.–2 Oct. 1948)* (Uzbekgosizdat, Tashkent), pp. 216–220.

(g) 'The real significance of results in dispersive analysis', *Proc. 2nd All-Union Conference on Mathematical Statistics (Tashkent, 27 Sept.–2 Oct. 1948)* (Uzbekgosizdat, Tashkent), pp. 240–268.

(h) 'On the disintegration of drops in a turbulent flow', *Dokl. Akad. Nauk SSSR* 66, 825–828.

(i) 'Absolute value', *BSE-2*, Vol. 1, 32.

(j) 'Jacques Hadamard', *BSE-2*, Vol. 1, 388.

(k) 'Additive quantities', *BSE-2*, Vol. 1, 394.

(l) 'Axiom', *BSE-2*, Vol. 1, 613–616.

(m) 'Axonometry', *BSE-2*, Vol. 1, 617.

1950

(a) 'Unbiased estimates', *Izv. Akad. Nauk SSSR Ser. Mat.* 14, 303–326.

(b) 'On the question of the determination of the coefficient of temperature conduction of soil', *Izv. Akad. Nauk SSSR Ser. Geog. Geofiz.* 14, no. 2, 97–98.

(c) 'Algebra in the secondary school', *BSE-2*, Vol. 2, 61–62.

(d) 'Algebraic expression', *BSE-2*, Vol. 2, 64.

(e) 'Algorithm', BSE-2, Vol. 2, 65.

(f) 'Euclid's algorithm', *BSE-2*, Vol. 2, 65–67.

(g) 'Aleksandr Danilovich Aleksandrov', *BSE-2*, Vol. 2, 83.

(h) 'Pavel Sergeevich Aleksandrov', *BSE-2*, Vol. 2, 84.

(i) 'Asymptote', *BSE-2*, Vol. 3, 238–239.

(j) 'Asymptotic expressions', *BSE-2*, Vol. 3, 239.

(k) 'Naum Il'ich Akhiezer', *BSE-2*, Vol. 3, 565.

(l) 'Stefan Banach', *BSE-2*, Vol. 4, 183.

(m) 'Nina Karlovna Bari', *BSE-2*, Vol. 4, 245.

(n) 'Sergei Natanovich Bernstein', *BSE-2*, Vol. 5, 52.

(o) 'Infinite numbers', *BSE-2*, Vol. 5, 66–67.

(p) (With V. F. Kagan) 'Infinitesimal numbers', *BSE-2*, Vol. 5, 67–71.

(q) (With B. N. Delone) 'Elements at infinity', *BSE-2*, Vol. 5, 2.

(r) 'Infinity (in mathematics)', *BSE-2*, Vol. 5, 73–74.

(s) 'Biharmonic functions', *BSE-2*, Vol. 5, 159.

(t) 'Bilinear form', *BSE-2*, Vol. 5, 167.

(u) 'The law of large numbers', *BSE-2*, Vol. 5, 538–540.

(v) 'An exposition of the foundations of Lebesgue's theory of measure', *Uspekhi Mat. Nauk* 5, no. 1, 211–213.

1951

(a) 'On the question of differentiability of transition probabilities in time-homogeneous Markov processes with countably many states', *Uch. Zap. Moskov. Gos. Univ.* 148 (Matematika 4), 53–59.

(b) 'A generalization of Poisson's formula to the case of a sample from a finite aggregate', *Uspekhi Mat. Nauk* 6, no. 3, 133–134.

(c) 'Ivan Georgievich Petrovskii', *Uspekhi Mat. Nauk* 6, no. 3, 161–164.

(d) 'A statistical acceptance control for an admissible number of defective products equal to zero', Leningrad, 1–24.

(e) (With S. A. Yanovskii) 'Luitzen Egbert Jan Brouwer', *BSE-2*, Vol. 6, 62.

(f) 'Variational series', *BSE-2*, Vol. 6, 641.

(g) (With S. A. Yanovskii) 'Hermann Weyl', *BSE-2*, Vol. 7, 106.

(h) 'Quantity', *BSE-2*, Vol. 7, 340–341.

(i) 'Probable error', *BSE-2*, Vol. 7, 507.

(j) 'Probability', *BSE-2*, Vol. 7, 508–510.

(k) (With T. I. Kozlov) 'Sampling method', *BSE-2*, Vol. 9, 417–418.

(l) (With A. N. Khovanskii) 'Introductory article and comments on the book: *N. I. Lobachevskii, Complete collected works*' (Gostekhizdat, Moscow–Leningrad), vol. 5, pp. 329–332; 342–348.

1952

(a) 'On the question of drag in the velocity profile under turbulent flow in tubes', *Dokl. Akad. Nauk SSSR* 84, 20–30.

(b) 'Gaussian distribution', *BSE-2*, Vol. 10, 275.

(c) 'Geodesic curvature', *BSE-2*, Vol. 10, 481.
(d) 'David Hilbert', *BSE-2*, Vol. 11, 370–371.
(e) 'Histogram', *BSE-2*, Vol. 11, 447.
(f) 'Boris Vladimirovich Gnedenko', *BSE-2*, Vol. 11, 545.
(g) 'Homeomorphism', *BSE-2*, Vol. 12, 21.
(h) 'Homotopy', *BSE-2*, Vol. 12, 35.
(i) 'Motion (in geometry)', *BSE-2*, Vol. 13, 447.
(j) 'Binomial', *BSE-2*, Vol. 13, 518.
(k) 'Real numbers', *BSE-2*, Vol. 13, 570.
(l) 'Division', *BSE-2*, Vol. 13, 628.
(m) 'Discreteness', *BSE-2*, Vol. 14, 435.
(n) 'Dispersion', *BSE-2*, Vol. 14, 438.
(o) 'Distributivity', *BSE-2*, Vol. 14, 479.
(p) 'Distributive operator', *BSE-2*, Vol. 14, 479.
(q) 'Differential (in mathematics)', *BSE-2*, Vol. 14, 498–499.
(r) (With B. P. Demidovich and V. V. Nemytskii) 'Differential equations', *BSE-2*, Vol. 14, 520–526.
(s) 'Fiducial probability', *BSE-2*, Vol. 14, 616.
(t) 'Confidence limits', *BSE-2*, Vol. 14, 617.
(u) (With I. G. Bashmakova and A. P. Yushkevich) 'Mathematical symbols', *BSE-2*, Vol. 17, 115–119.
(v) 'Value of a digit', *BSE-2*, Vol. 17, 135.
(w) (With V. I. Bityutskov) 'Isomorphism', *BSE-2*, Vol. 17, 478–479.
(x) 'Isotropic lines', *BSE-2*, Vol. 17, 509.
(y) 'Concrete number', *BSE-2*, Vol. 17, 557.
(z) 'Vasilii Grigor'evich Imshenetskii', *BSE-2*, Vol. 17, 607.
(A) *Mathematics as a profession: to help those entering higher educational institutes* (Soviet. Nauka).

1953
(a) 'On the concept of algorithm', *Uspekhi Mat. Nauk* 8, no. 4, 175–176.
(b) 'Some papers of recent years in the field of limit theorems in the theory of probability', *Vestnik Moskov. Univ.* 10, 29–38.
(c) 'On dynamical systems with an integral invariant on a torus', *Dokl. Akad. Nauk SSSR* 93, 763–766.
(d) 'Mathematical induction', *BSE-2*, Vol. 18, 146.
(e) (With V. I. Glivenko) 'Integral', *BSE-2*, Vol. 18, 250–253.
(f) 'Probability integral', *BSE-2*, Vol. 18, 253.
(g) 'Interpolation', *BSE-2*, Vol. 18, 304–305.
(h) 'Intuitionism', *BSE-2*, Vol. 18, 319.
(i) 'Elimination of unknowns', *BSE-2*, Vol. 18, 483.
(j) 'Experiment', *BSE-2*, Vol. 18, 604.
(k) 'Method of exhaustion', *BSE-2*, Vol. 19, 50–51.
(l) 'Quadrant', *BSE-2*, Vol. 20, 434.
(m) 'Compactum', *BSE-2*, Vol. 22, 282.
(n) 'Constant', *BSE-2*, Vol. 22, 416.
(o) 'Continuum', *BSE-2*, Vol. 22, 454–455.
(p) 'Coordinates', *BSE-2*, Vol. 22, 524–525.
(q) 'Correlation', *BSE-2*, Vol. 23, 55–58.

1954
(a) 'On the conservation of conditionally periodic motions for a small change in Hamilton's functions', *Dokl. Akad. Nauk SSSR* 98, 527–530.
(b) 'General theory of dynamical systems and classical mechanics', *Proc. Internat. Congress Math.* (Amsterdam, 1954), Vol. 1, 315–333.
(c) 'The line', *BSE-2*, Vol. 25, 167–170.
(d) 'The law of small numbers', *BSE-2*, Vol. 26, 168.
(e) 'Andrei Andreevich Markov', *BSE-2*, Vol. 26, 294.
(f) 'Mathematics', *BSE-2*, Vol. 26, 464–483.
(g) 'Mathematical statistics', *BSE-2*, Vol. 26, 485–490.
(h) 'Mathematical physics', *BSE-2*, Vol. 26, 490.
(i) 'Richard von Mises', *BSE-2*, Vol. 27, 414.
(j) 'Multidimensional space'. *BSE-2*, Vol. 27, 660.
(k) (With P. S. Aleksandrov) 'Set theory', *BSE-2*, Vol. 28, 14–17.
(l) (With S. V. Fomin) *Elements of the theory of functions and functional analysis. A course of lectures.* Part I, Metric and normed spaces (Izdat. Moskov. Univ. English edition: Graylock Press, Rochester, NY, 1957).
(m) Chief editor's preface to the Russian translation of: R. Peter, *Recursive functions* (Izdat. Inostr. Lit., Moscow, 3–10. Third English edition, Academic Press, New York–London, 1967).

1955
(a) 'Estimates of the minimal number of elements of ε-nets in distinct function classes and their application to the question of representability of functions of several variables by superpositions of functions of fewer variables', *Uspekhi Mat. Nauk* 10, no. 1, 192–194.
(b) 'Orientation', *BSE-2*, Vol. 31, 188–189.
(c) 'Foundations of geometry', *BSE-2*, Vol. 31, 296.
(d) (With L. A. Skornyakov) 'Surface', *BSE-2*, Vol. 33, 346–347.
(e) 'Ordered numbers', *BSE-2*, Vol. 34, 238.
(f) 'Acceptance statistical control', *BSE-2*, Vol. 34, 498–499.
(g) As (a), but in *Dokl. Akad. Nauk* 101, 192–194.

1956
(a) 'On the Skorokhod convergence', *Teor. Veroyatnost. i Primenen.* 1, 239–247; = *Theory Probab. Appl.* 1, 215–222.
(b) 'Two uniform limit theorems for sums of independent random variables', *Teor. Veroyatnost. i Primenen.* 1, 426–436; = *Theory Probab. Appl.* 1, 384–394.
(c) (With Yu. V. Prokhorov) 'Zufällige Funktionen und Grenzverteilungssätze', *Bericht über die Tagung Wahrscheinlichkeitsrechnung und mathematische Statistik* (Berlin), pp. 113–126.
(d) 'Some fundamental questions of approximate and exact representation of functions of one and several variables', *Proc. 3rd All-Union Math. Congress* (Moscow), vol. 2, 28–29.
(e) 'On the Shannon theory of information transmission in the case of continuous signals' [in English], *IEEE Trans. Inform. Theory*, Vol. I, T-2, 102–108.
(f) 'On the representation of continuous functions of several variables by superpositions of continuous functions of fewer variables', *Dokl. Akad. Nauk SSSR* 108, 179–182.
(g) 'On certain asymptotic characteristics of completely bounded metric spaces', *Dokl. Akad. Nauk SSSR* 108, 385–388.
(h) (With I. M. Gel'fand and A. M. Yaglom) 'On a general definition of the amount of information', *Dokl. Akad. Nauk SSSR* 111, 745–748.
(i) 'Theory of probability', *Mathematics, its content, methods, and significance* (Izdat. Akad. Nauk SSSR, Moscow), Vol. 2, pp. 252–284.
(j) (With S. B. Stechkin) 'S. M. Nikol'skii (on his 50th birthday)' *Uspekhi Mat. Nauk* 11, no. 2, 239–244.
(k) 'Evgenii Borisovich Slutskii', *BSE-2*, Vol. 39, 378.
(l) 'Nikolai Vasil'evich Smirnov', *BSE-2*, Vol. 39, 406.

1957
(a) 'Theory of the transmission of information', *Plenary Session of the USSR Academy of Sciences on scientific problems of automatization of production, 15–20 October 1956* (Izdat. Akad. Nauk SSSR, Moscow), pp. 66–99.
(b) 'On the representation of continuous functions of several variables by superpositions of continuous functions of one variable and addition', *Dokl. Akad. Nauk SSSR* 114, 953–956.
(c) 'On a justification of the theory of real numbers', *Mat. Prosveshchenie* 2, 169–173.

1958
(a) *Mathematics as a profession* (2nd ed., Izdat. Moskov. Univ.).
(b) (With I. M. Gel'fand and I. M. Yaglom) 'The amount of information and entropy for continuous distributions', *Proc. 3rd All-Union Math. Congress* (Izdat. Akad. Nauk SSSR, Moscow), Vol. 3, pp. 300–320.
(c) 'Sufficient statistics', *BSE-2*, Vol. 51, 106.
(d) 'Information', *BSE-2*, Vol. 51, 129–130.
(e) 'Cybernetics', *BSE-2*, Vol. 51, 149–151.
(f) 'A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces', *Dokl. Akad. Nauk SSSR* 119, 861–864.
(g) 'Sur les propriétés des fonctions de concentrations de M. P. Lévy', *Ann. Inst. H. Poincaré* 16, no. 1, 27–34.
(h) 'On linear dimensionality of topological vector spaces', *Dokl. Akad. Nauk SSSR* 120, 239–241.
(i) (With V. A. Uspenskii) 'On the definition of an algorithm', *Uspekhi Mat. Nauk* 13, no. 4, 3–28; = *AMS Transl.* (2) 17 (1961), 217–245.
(j) 'General theory of dynamics and classical mechanics' [in English], *Proc. Internat. Congr. Math. Edinburgh*, Vol. 1, 315.

1959
(a) 'Entropy per unit of time as a metric invariant of automorphisms', *Dokl. Akad. Nauk SSSR* 124, 754–755.

(b) (With V. M. Tikhomirov) 'ε-entropy and ε-capacity of sets in a function space', *Uspekhi Mat. Nauk* 14, no. 2, 3–86; = *AMS Transl.* (2) 17 (1961), 277–364.
(c) 'Transition of branching processes in diffusion and related problems of genetics', *Teor. Veroyatnost. i Primenen.* 4, 233–236.
(d) 'Remarks on papers by R. A. Minlos and V. V. Sazonov', *Teor. Veroyatnost. i Primenen.* 4, 237–239.
(e) 'Theory of probability', *Forty years of mathematics in the USSR* (Fizmatgiz, Moscow), Vol. 1, 781–795.
(f) Preface to the book: W. R. Ashby, *Introduction to cybernetics* (Chapman and Hall, London, 1956) (Izdat. Inostr. Lit., Moscow), pp. 5–8.

1960
(a) (With O. V. Sarmanov) 'The work of S. N. Bernstein on probability theory', *Teor. Veroyatnost. i Primenen.* 5, 215–221; = *Theory Probab. Appl.* 5, 197–203.
(b) 'On the $\phi^{(n)}$ classes of Forte and Blanc-Lapierre', *Teor. Veroyatnost. i Primenen.* 5, 373; = *Theory Probab. Appl.* 5, 337.
(c) 'Random functions of several variables, almost all realizations of which are periodic', *Teor. Veroyatnost. i Primenen.* 5, 374; = *Theory Probab. Appl.* 5, 338.
(d) (With B. V. Gnedenko, Yu. V. Prokhorov and O. V. Sarmanov) 'The work of N. V. Smirnov and mathematical statistics (on his 60th birthday)', *Teor. Veroyatnost. i Primenen.* 5, 436–440; = *Theory Probab. Appl.* 5, 397–401.
(e) (With Yu. A. Rozanov) 'On strong mixing conditions for stationary Gaussian processes', *Teor. Veroyatnost. i Primenen.* 5, 222–227; = *Theory Probab. Appl.* 5, 204–208.
(f) (With B. V. Gnedenko) 'Aleksandr Yakovlevich Khinchin (obituary)', *Uspekhi Mat. Nauk* 15, no. 4, 97–110; = *Russian Math. Surveys* 15, no. 4, 93–106.
(g) *Mathematics as a profession* (3rd ed., Izdat. Moskov. Univ., Moscow).
(h) (With S. V. Fomin) *Elements of the theory of functions and functional analysis.* Vol. 2, *Measure, Lebesgue integral, Hilbert space* (Izdat. Moskov. Univ., Moscow).

1961
(a) 'Automata and life' (Abstract of a paper read at the methodological seminar of the Faculty of Mechanics and Mathematics of Moscow State University of 5 April 1961), *Mash. Per. i Prikl. Lingvist.* 6, 3–8.
(b) 'Automata and life', *Tekhnika Molodezhi* no. 10, 16–19; no. 11, 30–33.
(c) 'Remark on a lecture by V. K. Lezerson', *Teor. Veroyatnost. i Primenen.* 6, 367.
(d) (With P. S. Aleksandrov) 'Properties of inequalities and the concept of approximate calculations. Irrational numbers', *Questions of the teaching of mathematics in the secondary school* (Uchpedgiz, Moscow).

1962
(a) (With M. Arato and Ya. G. Sinai) 'On the estimation of the parameters of a stationary Gaussian Markov process', *Dokl. Akad. Nauk SSSR* 146, 747–750; = *Soviet Math. Dokl.* 3, 1368–1371.
(b) (With A. M. Kondratov) 'The rhythmics of Mayakovsky's poems', *Vopr. Yazykoznaniya*, no. 3, 62–74.
(c) (With L. S. Pontryagin and E. F. Mishchenko) 'A probability problem in optimal control', *Dokl. Akad. Nauk SSSR* 145, 993–995; = *Soviet Math. Dokl.* 3, 1143–1145.
(d) 'The work of B. V. Gnedenko on probability theory', *Teor. Veroyatnost. i Primenen.* 7, 323–329.
(e) 'Précisions sur la structure locale de la turbulence dans un fluide visqueux aux nombres de Reynolds élévés' (Marseille, Aug.–Sept. 1961), *Proc. Coll. Internat. du Mécanique de la Turbulence* (Paris), pp. 447–458.
(f) 'A refinement of previous hypotheses concerning the local structure of turbulence of a viscous incompressible fluid at high Reynolds number' [in English], *J. Fluid Mech.* 13, no. 1, 82–85. See also *Proc. Internat. Colloq. 'Mech. de la Turbulence'* (CNRS, 1962), pp. 447–458.
(g) (With I. I. Gikhman and V. S. Korolyuk) 'B. V. Gnedenko (on his fiftieth birthday)', *Uspekhi Mat. Nauk* 17, no. 4, 191–200; = *Russian Math. Surveys* 17, no. 4, 105–113.
(h) (With S. A. Aivazyan, L. D. Meshalkin and V. F. Pisarenko) Review of the book: A. M. Dlin, *Mathematical statistics in technology*, Teor. Veroyatnost. i Primenen. 7, 243–248.

1963
(a) 'Approximation to distributions of sums of independent random variables by infinitely divisible distributions', *Trudy Moskov. Mat. Obshch.* 12, 437–451.
(b) 'Discrete automata and finite algorithms', *Proc. 4th All-Union Math. Congress* (Izdat. Leningrad. Univ., Leningrad), Vol. 1, 120.
(c) 'Various approaches to an estimate of the difficulty of an approximate definition and calculation of functions', *Proc. Internat. Congress Math., Stockholm*, pp. 369–376.

(d) 'On tables of random numbers' [in English], *Sankhyā Ser. A* 25, 369–376.
(e) 'The study of Mayakovskii's rhythmics', *Vopr. Yazykoznaniya* no. 4, 64–71.
(f) (With A. V. Prokhorov) 'On contemporary Russian poetry: general characteristics', *Vopr. Yazykoznaniya* no. 6, 84–95.
(g) (With A. V. Prokhorov) 'Statistics and probability theory in research into Russian poetry', *Proc. Symp. on complex investigation of artistic creation* (Nauka, Leningrad), p. 23.
(h) 'How I became a mathematician', *Ogonëk* no. 48, 12–13.
(i) Preface to the book: C. E. Shannon, *Papers on information theory and cybernetics* (Izdat. Inostr. Lit., Moscow).

1964
(a) (With A. V. Prokhorov) 'On contemporary dol'nik Russian poetry: Statistical characteristics of dol'nik in the works of Mayakovskii, Bagritskii, and Akhmatova', [Dol'nik is a form of tonic verse that has a trisyllabic metre and allows the omission of one or two unstressed syllables in each line], *Vopr. Yazykoznaniya*, no. 1, 75–94.
(b) 'On the metre of Pushkin's "Songs of the Western Slavs"', *Russkaya Lit.* no. 1, 98–111.
(c) Preface to the Russian translation of the book: W. Feller, *An introduction to probability theory and its applications* (2nd ed., Wiley, New York, 1957) (Mir, Moscow), pp. 5–6.

1965
(a) 'Three approaches to the definition of the concept "quantity of information"', *Probl. Peredachi Informatsii* 1, no. 1, 3–11.
(b) 'The volume of knowledge about mathematics for eight year olds in school (commission on mathematical education of the Math. Division of the USSR Academy of Sciences)', *Mat. v Shkole* no. 2, 21–24.
(c) 'Geometrical mappings in a school geometry course', *Mat. v Shkole* no. 2, 24–29.
(d) (With I. M. Yaglom) 'On the content of a school mathematics course', *Mat. v Shkole* no. 4, 53–62.
(e) 'Functions, graphs, continuous functions', *Mat. v Shkole* no. 6, 12–21.
(f) 'Remarks about the analysis of the rhythm of Mayakovsky's "Lines on a Soviet Passport"', *Vopr. Yazykoznaniya* no. 3, 70–75.
(g) 'Natural numbers and positive scalar quantities', *School Mathematics. Lectures and Problems*, no. 4–5 (Izdat. Moskov. Univ., Moscow), pp. 19–35.

1966
(a) *Introduction to analysis* (Izdat. Moskov. Univ., Moscow).
(b) 'P. S. Aleksandrov and the theory of $\delta\sigma$-operations', *Uspekhi Mat. Nauk* 21, no. 4, 275–278.
(c) 'Textbooks for the academic year 1966/67', *Mat. v Shkole* no. 2, 26–30.
(d) 'A school definition of identity', *Mat. v Shkole* no. 2, 33–35.
(e) 'Introduction to S. B. Suvorova's article "An attempt at an earlier introduction of the elements of differential calculus"', *Mat. v Shkole* no. 4, 23.
(f) 'Textbooks for the academic year 1966/67', *Mat. v Shkole* no. 6, 31–37.
(g) 'Geometry on the sphere and geology', *Nauka i Zhizn'* no. 2, 32.
(h) 'A problem from the theory of curves', *School Mathematics. Lectures and Problems*, no. 8, 35.

1967
(a) (With Ya. M. Barzdin') 'On the realization of networks in three-dimensional space', *Problemy Kibernet.* 19, 261–268.
(b) 'Textbooks for the academic year 1966/67', *Mat. v Shkole* no. 1, 43–48.
(c) 'Optional studies in mathematics', *Mat v Shkole* no. 2, 2–3.
(d) 'New programmes and some basic problems in perfecting a mathematics course in the secondary school', *Mat. v Shkole* no. 2, 4–13.
(e) 'Content of optional studies in mathematics in the academic years 1967/68 and 1968/69', *Mat. v Shkole* no. 2, 33–38.
(f) 'Programmes of special courses in mathematics', *Mat. v Shkole* no. 3, 73–75.
(g) 'Programmes of special courses in mathematics', *Mat. v Shkole* no. 4, 58–59.
(h) 'Changes in the text of A. N. Barsukov's algebra textbook for the 6th to 8th classes', *Mat. v Shkole* no. 6, 22–24.

1968
(a) 'Some theorems on algorithmic entropy and algorithmic amount of information', *Uspekhi Mat. Nauk* 21, no. 2, 201.
(b) 'Design of a programme in mathematics for the secondary school', *Mat. v Shkole* no. 1, 4–23.
(c) 'Generalization of the concept of a power and the exponential function', *Mat. v Shkole* no. 1, 24–32.
(d) 'A programme in mathematics for the secondary school', *Mat. v Shkole* no. 2, 5–20.

(e) 'On studying the exponential function and logarithms in primary school', *Mat. v Shkole* no. 2, 23–25.
(f) 'On new programmes in mathematics', *Mat. v Shkole* no. 2, 21–22.
(g) 'Introduction to the theory of probability and combinatorics', *Mat. v Shkole* no. 2, 63–72.
(h) Appendix to a review of Yu. A. Shikhanovich (of a book by A. A. Stolyar), *Mat. v Shkole* no. 3, 92.
(i) *Elements of the theory of functions and functional analysis* (2nd revised ed., Fizmatgiz, Moscow; English transl. of 1st ed., Graylock Press, Rochester, NY, 1957).
(j) (With A. V. Prokhorov) 'On the basis of the classical Russian metre', *Commonwealth of science and mysteries of creation* (Iskusstvo, Moscow), pp. 397–432.
(k) 'An example of the study of metre and its metric variants', *Theory of verse* (Nauka, Leningrad), pp. 145–167.

1969
(a) 'On the logical foundations of information theory and probability theory', *Probl. Peredachi Informatsii* 5, no. 3, 3–7.
(b) 'Sergei Natanovich Bernstein (obituary)', *Uspekhi Mat. Nauk* 24, no. 3, 211–218; = *Russian Math. Surveys* 24, no. 3, 169–176.
(c) Letter to the Editor (on errors in an article by B. E. Veits), *Mat. v Shkole* no. 2, 93.
(d) 'Scientific foundations for a school mathematics course. First lecture. Modern views on the nature of mathematics', *Mat. v Shkole* no. 3, 12–17.
(e) 'Scientific foundations for a school mathematics course. Second lecture. Natural numbers', *Mat. v Shkole* no. 5, 8–17.
(f) 'New developments in school mathematics', *Nauka i Zhizn'* no. 3, 62–66.

1970
(a) Viktor Nikolaevich Zasukhin (in memory of Soviet mathematicians who fell in the Second World War)', *Uspekhi Mat. Nauk* 25, no. 3, 243; = *Russian Math. Surveys* 25, no. 3, 179.
(b) 'Gleb Aleksandrovich Seliverstov (obituary)', *Uspekhi Mat. Nauk* 25, no. 3, 244–245; = *Russian Math. Surveys* 25, no. 3, 180–181.
(c) 'Scientific foundations for a school mathematics course. Third lecture. Generalization of the concept of number. Non-negative rational numbers', *Mat. v Shkole* no. 4, 27–32.
(d) (With A. F. Semenovich) 'On an experimental textbook of geometry for the sixth class', *Mat. v Shkole* no. 4, 21–34.
(e) (With A. F. Semenovich and R. S. Cherkasov) 'Teaching materials on geometry for the fifth class', *Mat. v Shkole* no. 5, 30–45.
(f) 'What is a function?', *Kvant* no. 1, 27–36.
(g) 'Problem no. 3', *Kvant* no. 1, 52–53.
(h) 'What is the graph of a function?', *Kvant* no. 2, 3–13.
(i) 'Tiling by regular polygons', *Kvant* no. 3, 24.
(j) (With F. L. Varpakhovskii) 'On the solution of Hilbert's 10th problem', *Kvant* no. 7, 39–44.
(k) Preface to the article: V. G. Boltyanskii and N. Kh. Rozov, 'Lenin's theory of knowledge and mathematical concepts', *Kvant* no. 7, 2.
(l) (With A. F. Semenovich and R. S. Cherkasov) *Geometry in the sixth class, an experimental textbook* (Prosveshchenie, Moscow).

1971
(a) 'Quantity', *BSE-3*, Vol. 4, 456–457.
(b) 'Norbert Wiener', *BSE-3*, Vol. 5, 72.
(c) 'David Hilbert', *BSE-3*, Vol. 6, 519.
(d) 'A system of basic concepts and notation for a school mathematics course', *Mat. v Shkole* no. 2, 17–22.
(e) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) 'From an experimental textbook of geometry for the seventh class', *Mat. v Shkole* no. 3, 9–17.
(f) 'Review of the lecture "Elements of logic in the modern school"' (at the Commission on Mathematics of the Academic Methods Council of the USSR Ministry of Education), *Mat. v Shkole* no. 3, 91.
(g) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) 'On the new edition of an experimental textbook of geometry for the sixth class', *Mat. v Shkole* no. 4, 23–25.
(h) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) 'On the new edition of the experimental textbook on geometry for the sixth class', *Mat. v Shkole* no. 5, 25–38.
(i) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) 'From a new experimental textbook on geometry for the sixth class' (Geometrical constructions), *Mat. v Shkole* no. 6, 13–21.
(j) 'Modern mathematics and mathematics in the secondary school', *Mat. v Shkole* no. 6, 2–3.
(k) (With V. A. Gusev, A. V. Sisinskii and A. A. Shershevskii) *A mathematics course for schools of physics and mathematics* (Izdat. Moskov. Univ., Moscow).
(l) 'Summer school on Lake Ruby', Prosveshchenie, Moscow.

(m) A letter to P. L. Kapitsa, *Vopr. Phil.* no. 9.
(n) (With other authors) 'Naum Il'ich Akhiezer (on his seventieth birthday)', *Uspekhi Mat. Nauk* 26, no. 6, 257–261; = *Russian Math. Surveys* 26, no. 6, 233–237.
(o) (With other authors) 'Ivan Georgievich Petrovskii (on his seventieth birthday)', *Uspekhi Mat. Nauk* 26, no. 2, 3–24; = *Russian Math. Surveys* 26, no. 2, 3–28.
(p) (With A. F. Semenovish, F. F. Nagibin and R. S. Cherkasov) *Geometry for the sixth class, a textbook* (Prosveshchenie, Moscow).
(q) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) *Geometry for the seventh class, an experimental texbook* (Prosveshchenie, Moscow).

1972
(a) 'The complexity of specification and construction of mathematical objects', *Uspekhi Mat. Nauk* 27, no. 2, 159.
(b) 'Integral', *BSE-3*, Vol. 10, 586.
(c) 'The method of exhaustion', *BSE-3*, Vol. 10, 586.
(d) (With S. V. Fomin) *Elements of the theory of functions and functional analysis* (3rd revised ed., Fizmatgiz, Moscow).
(e) 'The supervisor of studies', in L. Neiman, *The joy of discovery* (Detskaya Literatura, Moscow).
(f) 'Qualitative study of mathematical models of population dynamics', *Probl. Kibernet.* 35, no. 2, 101–106.
(g) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) 'From a new textbook on geometry for the sixth class', *Mat. v Shkole* no. 1, 22–31.
(h) (With R. S. Cherkasov) 'Boris Vladimirovich Gnedenko', *Mat. v Shkole* no. 1, 85–86.
(i) 'On a letter of N. Ya. Vilenkin', *Mat. v Shkole* no. 6, 34–35. (Vilenkin's letter "Equality or congruence" is in the same issue.)
(j) 'Agreement of teaching of mathematics and physics', *Proc. All-Union Conf. on the problem of educational work in schools and classes with a deep study of separate subjects* (*Sci. Research Inst., USSR Acad. Pedagog. Sci.*).
(k) (With other authors) 'Georgii Pavlovich Tolstov (on his sixtieth birthday)', *Uspekhi Mat. Nauk* 27, no. 1, 255–264; = *Russian Math. Surveys* 27, no. 1, 207–217.
(l) (With P. S. Aleksandrov and P. L. Ul'yanov) 'Dmitrii Evgen'evich Men'shov (on his eightieth birthday)', *Uspekhi Mat. Nauk* 27, no. 2, 185–195; = *Russian Math. Surveys* 27, no. 2, 161–171.
(m) (With Yu. K. Belyaev and A. D. Solov'ev) 'Boris Vladimirovich Gnedenko (on his sixtieth birthday)', *Uspekhi Mat. Nauk* 27, no. 2, 197–202; = *Russian Math. Surveys* 27, no. 2, 173–179.
(n) (With other authors) 'Leonid Vital'evich Kantorovich (on his sixtieth birthday)', *Uspekhi Mat. Nauk* 27, no. 3, 221–227; = *Russian Math. Surveys* 27, no. 3, 193–201.
(o) (With other authors) 'Georgii Fedorovich Rybkin (obituary)', *Uspekhi Mat. Nauk* 27, no. 5, 223–225; = *Russian Math. Surveys* 27, no. 5, 165–167.
(p) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) *Geometry for the eighth class, an experimental textbook* (Prosveshchenie, Moscow).
(q) (With A. F. Semenovich, F. F. Nagibin and R. S. Cherkasov) *Geometry for the sixth class* (2nd ed., Prosveshchenie, Moscow).
(r) 'Teachers cannot be replaced', *Komsomol'skaya Pravda*, 15 January, 2.

1973
(a) 'The continuum', *BSE-3*, Vol. 13, 64.
(b) 'The semilogarithmic and logarithmic network', *Kvant* no. 13, 64.
(c) 'Mathematics as a profession', *Kvant* no. 4, 12.
(d) (With A. F. Semenovich and R. S. Cherkasov) 'On the methodology of studying the topic "Parallelism and parallel shift" in the geometry course for the seventh class', *Mat. v Shkole* no. 1, 24–29.
(e) (With A. F. Semenovich and R. S. Cherkasov) 'On the structure of a new textbook on geometry for the seventh class', *Mat. v Shkole* no. 2, 17–29.
(f) (With P. S. Aleksandrov and O. A. Oleinik) 'Ivan Georgievich Petrovskii', *Mat. v Shkole* no. 4, 81–86.
(g) (With B. E. Veits and I. T. Demidov) 'Methodological remarks on an experimental textbook *Algebra and the elements of analysis for the ninth class*', *Mat. v Shkole* no. 5, 64.
(h) *Scientific foundations for a school mathematics course*, in the series: Programmes of pedagogical institutes (Prosveshchenie, Moscow).
(i) 'Materials for discussion at the Commission of School Terminology and Notation', (Educ. Method. Council, USSR Ministry of Education).

1974
(a) *Fundamental concepts of probability theory* (2nd ed., Nauka, Moscow). German translation: *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer-Verlag, Berlin-Heidelberg-New York, 1973).

(b) (With P. S. Aleksandrov and O. A. Oleinik) 'In memoriam Ivan Georgievich Petrovskii (18 January 1901–15 January 1974)', *Trudy Moskov. Mat. Obshch.* 32, 5–10; = *Trans. Moscow Math. Soc.* 31, 1–12.
(c) 'Ivan Georgievich Petrovskii', *Uspekhi Mat. Nauk* 29, no. 2, 3–5; = *Russian Math. Surveys* 29, no. 2, 3–5.
(d) 'Andrei Andreevich Markov', *BSE-3*, Vol. 15, 379.
(e) 'Mathematics', *BSE-3*, Vol. 15, 467–478.
(f) (With Yu. V. Prokhorov) 'Mathematical statistics', *BSE-3*, Vol. 15, 480–484.
(g) 'Multidimensional space', *BSE-3*, Vol. 16, 372.
(h) 'Orientation', *BSE-3*, Vol. 18, 509–510.
(i) (With I. T. Tropin and K. V. Chernishev) 'Bringing up a new generation', *Vestnik Vysshei Shkoly* no. 6, 26–33.
(j) 'New programmes: special schools', *Modern mathematical education* (Prosveshchenie, Moscow), pp. 5–12.
(k) 'A boarding school for the university. What is it for?', *Mat. v Shkole* no. 2, 56–60 (in the section 'Ten years of the school of physics and mathematics at Moscow State University').
(l) (With A. F. Semenovich) 'Anna Maksimilianovna Fisher (obituary)', *Mat. v Shkole* no. 2, 87.
(m) 'The sieve of Eratosthenes', *Kvant* no. 10, 2.

1975
(a) (With Yu. K. Belyaev) 'Acceptance statistical control', *BSE-3*, Vol. 20, 572–573.
(b) (With S. I. Shvartsburd) 'Algebra and the elements of analysis. The method of mathematical induction', *Mat. v Shkole* no. 1, 8–14.
(c) 'The elements of combinatorics', *Mat. v Shkole* no. 2, 16–25.
(d) (With O. S. Ivashev-Musatov) 'Real numbers, infinite sequence and their limits', *Mat. v Shkole* no. 2, 25–35.
(e) (With V. K. Dzyadyk and L. D. Kudryavtsev) 'Sergei Mikhailovich Nikol'skii (on his seventieth birthday)', *Uspekhi Mat. Nauk* 30, no. 4, 271–280; = *Russian Math. Surveys* 30, no. 4, 193–202.
(f) (With P. S. Aleksandrov and S. L. Sobolev) 'Ol'ga Arsen'evna Oleinik', *Vestnik Moskov. Univ. Ser. Mat. Mekh.* 4, 119–124.

1976
(a) (With S. V. Fomin) *Elements of the theory of functions and functional analysis* (4th revised ed., Nauka, Moscow. English translation of 1st ed.: Graylock Press, Albany, NY, 1961).
(b) (With S. I. Shvartsburd) 'Trigonometric functions, their graphs and derivatives, in the textbook for the tenth class', *Mat. v Shkole* no. 1, 10–25.
(c) (With G. A. Gal'perin) 'The 38th Moscow Mathematical Olympiad (February–March 1975)', *Mat. v Shkole* no. 4, 68–72.
(d) 'The integral in the textbook for the tenth class', *Mat. v Shkole* no. 6, 15–17.
(e) 'Groups of transformations', *Kvant* no. 10, 2–5.
(f) (With A. V. Arkhangel'skii, A. A. Mal'tsev and O. A. Oleinik) 'Pavel Sergeevich Aleksandrov (on his eightieth birthday)', *Uspekhi Mat. Nauk* 31, no. 5, 3–15; = *Russian Math. Surveys* 31, no. 5, 1–13.
(g) (With P. S. Aleksandrov, B. V. Gnedenko and Yu. V. Prokhorov) 'Tashmukhamed Alievich Sarymsakov (on his sixtieth birthday)', *Uspekhi Mat. Nauk* 31, no. 2, 241–246; = *Russian Math. Surveys* 31, no. 2, 215–221.

1977
(a) 'Infinity', *Mat. Entsiklopediya* Vol. 1, 455–458.
(b) 'Quantity', *Mat. Entsiklopediya* Vol. 1, 651–653.
(c) 'Probability', *Mat. Entsiklopediya* Vol. 1, 667–669.
(d) (With V. V. Vavilov) 'The school of physics and mathematics at Moscow State University', *Kvant* no. 1, 56–57.
(e) (With other authors) 'Adol'f Pavlovich Yushkevich (on his seventieth birthday)', *Uspekhi Mat. Nauk* 32, no. 3, 197–202; = *Russian Math. Surveys* 32, no. 3, 145–153.

1978
(a) 'What is a function?', *Mat. v Shkole* no. 2, 27–29 (in connection with G. V. Dorofeev's article 'The concept of a function in mathematics and at school' in the same issue).
(b) 'On introducing the concepts of dialectical materialism in mathematics and physics lessons', *Mat. v Shkole* no. 3, 6–9.
(c) 'Design of a programme in mathematics for the secondary school', *Mat. v Shkole* no. 4, 7–32.
(d) (With O. A. Oleinik) 'Sergei L'vovich Sobolev (on his seventieth birthday)', *Mat. v Shkole* no. 6, 67–73.
(e) (With A. M. Abramov) 'New programmes in the French secondary school', *Mat. v Shkole* no. 6, 74–78.

(f) 'How I became a mathematician. What is mathematics? Science as your profession', *Znanie* no. 11, 5–9.
(g) (With I. G. Zhurbenko) 'Estimates of spectral functions of stochastic processes', Lecture to the 11th European Conf. on Statistics, Oslo, 14–18 August 1978.
(h) 'On the formation of a dialectical-materialistic view of the world in schoolchildren in mathematics and physics lessions', *The role of educational literature in the formation of a view of the world in schoolchildren* (Pedagogika, Moscow), pp. 69–74.
(i) (With other authors) 'Mark Gregor'evich Krein (on his seventieth birthday)' *Uspekhi Mat. Nauk* 33, no. 3, 197–203; = *Russian Math. Surveys* 33, no. 3, 185–193.

1979
(a) (With A. V. Bulinskii) 'Linear sampling estimates of sums', *Teor. Veroyatnost. i Primenen.* 24, 241–251; = *Theory Probab. Appl.* 24, 241–252.
(b) (With V. V. Vavilov and I. T. Tropin) 'The first 20 years of the school of mathematics and physics at Moscow State University', *Kvant* no. 1, 55–57.
(c) (With A. F. Semenovich and R. S. Cherkasov) 'On the textbook *Geometry 6–8*', *Mat. v Shkole* no. 3, 38–42.
(d) (With V. D. Belousov, V. G. Boltyanskii et al.) 'Aleksei Ivanovich Markushevich (obituary)', *Mat. v Shkole* no. 6, 77–78.
(e) (With A. M. Abramov, O. S. Ivashev-Musatov, B. M. Ivlev and S. I. Shvartsburd) 'The exponential and logarithmic functions', *Mat. v Shkole* no. 6, 22–27.

1980
(a) 'A dialectical-materialistic view of the world in school courses of mathematics and physics', *Kvant* no. 4, 15–18.
(b) (With A. M. Abramov, O. S. Ivashev-Musatov, B. M. Ivlev and S. I. Shvartsburd) 'On the textbook *Algebra and analysis for the 9th and 10th classes*', *Mat. v Shkole* no. 6, 22–27.

1981
(a) (With S. V. Fomin) *Elements of the theory of functions and functional analysis* (5th ed., Nauka, Moscow).
(b) (With V. V. Vavilov and I. T. Tropin) *The school of physics and mathematics at Moscow State University*, Mathematics and cybernetics, no. 5 (Izdat. Moskov. Univ., Moscow).
(c) (With A. F. Semenovich and R. S. Cherkasov) *Geometry for the 6th and 8th classes*: textbook (3rd ed., Prosveshchenie, Moscow).
(d) 'On the concept of a vector in the mathematics course of the secondary school', *Mat. v Shkole* no. 3, 7–8.
(e) (With A. M. Abramov) 'On the question of taking the first steps in the topic "Vectors"', *Mat. v Shkole* no. 3, 8–11.
(f) 'Review of L. S. Pontryagin's book *Infinitesimal analysis*', *Mat. v Shkole* no. 5, 73–74.
(g) (With P. S. Aleksandrov, B. V. Gnedenko, S. S. Demidov, S. S. Petrova, K. A. Rybnikov and A. P. Yushkevich) 'Izabella Grigor'evna Bashmakova (on her sixtieth birthday)', *Mat. v Shkole* no. 1, 73–74.
(h) (With other authors) 'Vladimir Mikhailovich Alekseev (obituary)', *Uspekhi Mat. Nauk* 36, no. 4, 177–182; = *Russian Math. Surveys* 36, no. 4, 201–206.
(i) (With other authors) 'Naum Il'ich Akhiezer (obituary)', *Uspekhi Mat. Nauk* 36, no. 4, 183–184; = *Russian Math. Surveys* 36, no. 4, 207–208.
(j) (With other authors) 'Izabella Grigor'evna Bashmakova (on her sixtieth birthday)', *Uspekhi Mat. Nauk* 36, no. 5, 211–214; = *Russian Math. Surveys* 36, no. 5, 187–190.
(k) (With other authors) 'In memory of Mikhail Alekseevich Lavrent'ev', *Uspekhi Mat. Nauk* 36, no. 2, 3–10; = *Russian Math. Surveys* 36, no. 2, 1–10.
(l) (With other authors) 'In memory of Anatolii Illarionovich Shirshov', *Uspekhi Mat. Nauk* 36, no. 5, 153–157; = *Russian Math. Surveys* 36, no. 5, 129–133.
(m) (With other authors) 'Sagdy Khasonovich Sirazhdinov (on his sixtieth birthday)', *Uspekhi Mat. Nauk* 36, no. 6, 237–242; = *Russian Math. Surveys* 36, no. 6, 208–214.
(n) (With other authors) 'Aleksandr Filippovich Timan (on his sixtieth birthday)', *Uspekhi Mat. Nauk* 36, no. 2, 221–225; = *Russian Math. Surveys* 36, no. 2, 213–218.
(o) (With V. V. Vavilov and I. T. Tropin) *The school of physics and mathematics at Moscow State University* (Znanie, Moscow).
(p) (With P. S. Aleksandrov) 'A. I. Markushevich as a historian of mathematics' [in English], *Historia Math.* 8, 125–132.

1982
(a) (With A. G. Dragalin) *Introduction to mathematical logic* (Izdat. Moskov. Univ., Moscow).

(b) (With I. G. Zhurbenko and A. V. Prokhorov) *Introduction to probability theory*, Library "Kvant", no. 23 (Nauka, Moscow).
(c) (With Yu. V. Prokhorov) 'Mathematics', *Mat. Entsiklopediya*, Vol. 3, 576–581.
(d) (With R. S. Cherkasov) 'Boris Vladimirovich Gnedenko', *Mat. v Shkole* no. 1, 72–73.
(e) (With V. A. Zalgaller) 'Leonid Vital'evich Kantorovich (on his seventieth birthday)', *Mat. v Shkole* no. 2, 77–78.
(f) 'On the concept of limit in the general-educational school', *Mat. v Shkole* no. 5, 56.
(g) 'Newton and modern mathematical thinking', *Mat. v Shkole* no. 6, 58.
(h) (With Yu. K. Belyaev and A. D. Solov'ev) 'Boris Vladimirovich Gnedenko (on his seventieth birthday)', *Uspekhi Mat. Nauk* 37, no. 6, 243–248; = *Russian Math. Surveys* 37, no. 6, 275–281.
(i) (With S. M. Nikol'skii, V. A. Skvortsov and P. L. Ul'yanov) 'Dmitrii Evgen'evich Men'shov (on his ninetieth birthday)', *Uspekhi Mat. Nauk* 37, no. 5, 209–219; = *Russian Math. Surveys* 37, no. 5, 203–215.

1983
(a) (With A. M. Abramov, B. E. Veits, O. S. Ivashev-Musatov and S. I. Shvartsburd) *Algebra and the elements of analysis: textbook for the 9th and 10th classes of the secondary school,* (4th ed., Prosveshchenie, Moscow).
(b) (With B. V. Gnedenko) 'Pavel Sergeevich Aleksandrov', *Mat. v Shkole* no. 1, 47–48.
(c) 'On the textbook *Geometry* by A. V. Pogorelov', *Mat. v Shkole* no. 2, 45.
(d) 'Combinatorial foundations of information theory and the calculus of probabilities', *Uspekhi Mat. Nauk* 38, no. 4, 27–36; = *Russian Math. Surveys* 38, no. 4, 29–40.

1984
(a) 'Remarks on the concept of a set in the school mathematics course', *Mat. v Shkole* no. 1, 52–53.
(b) (With O. A. Oleinik) 'S. L. Sobolev and modern mathematics', *Mat. v Shkole* no. 1, 73–77.
(c) 'Analysis of the metrical structure of Pushkin's poem "Arion"', *Problems of the theory of poetry* (Nauka, Leningrad), pp. 118–120.

1985
(a) 'A model of the rhythmic structure of Russian speech, adapted to the study of the metric of classical Russian poetry', *Russian verse: traditions and problems of development* (Nauka, Moscow), pp. 113–134.
(b) *Collected works. Mathematics and mechanics* (Nauka, Moscow).
(c) 'A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces', New edition, *Trudy Mat. Inst. Akad. Nauk* 169, no. 1, 94–98.

1986
(a) 'On scalar quantities', *Mat. v Shkole* no. 3, 32–33.
(b) 'Memories of P. S. Aleksandrov', *Uspekhi Mat. Nauk* 41, no. 6, 187–203; = *Russian Math. Surveys* 41, no. 6, 225–246.
(c) *Theory of probability and mathematical statistics* (Nauka, Moscow).
(d) Preface to the book: J. Bernoulli, *On the law of large numbers* (Nauka, Moscow).

1987
(a) *Information theory and the theory of algorithms* (Nauka, Moscow).
(b) Speech of welcome to the participants of the first World Congress of the Bernoulli Society, *Teor. Veroyatnost. i Primenen*, 32, 218; = *Theory Probab. Appl.* 32, 200.
(c) (With V. A. Uspenskii) 'Algorithms and chance', *Teor. Veroyatnost. i Primenen.* 32, 425–455.

1988
(a) (With G. A. Gal'perin) *Mathematics – science and profession*, Library "Kvant", no. 64 (Nauka, Moscow).
(b) (With Yu. V. Prokhorov and A. N. Shiryaev) 'Probabilistic-statistical methods of discovering spontaneously arising effects', *Trudy Mat. Inst. Akad. Nauk*.

Comments on the Kolmogorov bibliography

The main source used was the list prepared by Professor A. N. Shiryaev, Kolmogorov's literary executor, which is referred to in what follows as 'Complete list'. This was compared with the list of Kolmogorov's works as published in *Uspekhi Mat. Nauk* 18, no. 5 (1963), 28, no. 5 (1973), and 38, no. 4 (1983), and translated in

Russian Math. Surveys. The first part of the list, published in *Uspekhi Mat. Nauk*, 8 (1953), was before the days of *Russian Math. Surveys*, so for the early part of the book *Matematika v SSSR za sorok let 1917–1957* (Mathematics in the USSR for the 40 years 1917–1957) was used.

In the notes below, the following abbreviations will be used: *UMN* for *Uspekhi Mat. Nauk*, *SSSR* for the book referred to above, and *MR* for *Math. Reviews*.

[1923a]. *SSSR* gives page numbers 324–326. Here and later there may be a discrepancy between different editions.
[1925c]. The title in *SSSR* is an abbreviation of this.
[1925d]. *SSSR* gives page numbers 23–28.
[1927a]. *SSSR* gives page numbers 919–921.
[1933c]. The complete list gives the German double s, but *SSSR* gives the ordinary ss.
[1934d]. *SSSR* gives page numbers 291–295.
[1936j]. *SSSR* gives page numbers 847–850.
[1937e]. *SSSR* gives page numbers 355–360.
[1938c]. *SSSR* gives page numbers 359–401.
[1938p]. There is a little uncertainty as to whether this is the correct work of Lebesgue.
[1941a]. *SSSR* gives page numbers 1–40, the complete list gives 1–10. According to *MR*, this issue consists of 40 pages.
[1943b]. Presumably this is one of the articles in [1943a].
[1946d]. *SSSR* gives page numbers 27–42.
[1949c]. The title on the complete list makes it look like 'Word-formation'. In fact, it is 'stratification', as in *SSSR* and confirmed by *MR*.
[1952q]. *SSSR* inserts the word 'in mathematics', presumably to distinguish it from differential in engineering.
[1963d]. This journal is usually known as *Sankhyā*, though it is sometimes called the *Indian Journal of Statistics*, as on the complete list.
[1965a]. It is Volume 1, as in *UMN* and *MR*.
[1966b]. The complete list gives δ_s, *UMN* gives δs, and *MR* gives $\delta\sigma$.
[1970d]. Compare [1970l]. Probably (d) is a summary of (l).
[1971e]. Compare [1971q]. Probably (e) is a summary of (q).
[1971l]. This must be Lake Ruby in the USA. Presumably Kolmogorov went there, and this is his report.
[1972g]. Compare [1972q].
[1972h]. Compare [1972m].
[1972j]. It has proved impossible to trace the acronym SIMO.
[1979b]. The complete list says 20 years, *UMN* says 15 years!
[1981b]. Compare [1981o].
[1982d]. Compare [1982h].

Statistical Laboratory
16 Mill Lane
Cambridge CB2 1SB