Statistical Science
1991, Vol. 6, No. 3, 303–312

Kolmogorov as I Remember Him

David G. Kendall

1. EARLY YEARS

Andrei Nikolaevitch Kolmogorov was born in 1903 during a journey from the Crimea to his mother's home. He was the son of parents not formally married. His mother, Mariya Yakovlevna Kolmogorova, died in childbirth, and her son 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 the love of a mother, and Andrei Nikolaevitch responded with the love of a son. It is warming to be able to record that she lived until 1950, and so was able to witness some of his greatest achievements.

Andrei Nikolaevitch is always known to us by the family name of his maternal grandfather Yakov Stepanovitch Kolmogorov, and it was in the Kolmogorov family home at Tunoshna that he spent his earliest years. During his childhood the family home housed a clandestine printing press, and family traditions record that compromising documents were sometimes hidden under his cradle.

Of Kolmogorov's father, Nikolai Kataev, we know 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.

Kolmogorov went to Moscow in 1920 as a student of mathematics, but he also attended lectures in metallurgy. In addition to this he took part in a seminar on Russian history, where he presented the results of his first piece of research, on "Landholding in Novgorod in the 15th and 16th centuries."

We are told how this was received by his professor: "You have supplied one proof of your thesis, and in mathematics this would perhaps suffice, but we historians prefer to have ten proofs." This anecdote is usually told as a joke, but to those who know something about the limitations of such archives it will seem a fair comment. However, it is also on record that an expedition to the region later confirmed Kolmogorov's conjecture about the way in which the upper Pinega was settled.

A number of mathematicians stimulated Kolmogorov's earliest mathematical research, but perhaps his principal teacher was Stepanov. In 1922 Kolmogorov produced a synthesis of the French and Russian work on the descriptive theory of sets of points, and at about the same time he was introduced to Fourier series in Stepanov's seminar. This was when he made his first mathematical discovery —*that there is no such thing as a slowest possible rate of convergence to zero for the Fourier cosine coefficients of an integrable function.*

Nearly 30 years ago I gave a lecture in Tbilisi in which I proved that in the transient aperiodic case the diagonal Markov transition probabilities $p_{ii}^{(n)}$ for fixed i always form a sequence of Fourier cosine coefficients, and I remarked that it would be interesting to see what one could deduce from this fact concerning the rate of convergence to zero as n tends to infinity. Kolmogorov made a comment that I did not understand at the time because of language difficulties. It occurs to me now that he must have thinking of an application of his own first paper in this new context. Thus one can ask whether the sequence of such diagonal Markov transition probabilities also has no slowest rate of convergence to zero. I do not know the answer to that question.

Figure 1 shows Kolmogorov (wearing spectacles and leaning over to his left) at the Tbilisi meeting. Also in the picture are Dynkin and Gnedenko (to Kolmogorov's right) and many other well-known probabilists. As I don't myself appear in this photograph of the audience, I like to think that it may have been my lecture they were listening to! Recently I attended a lecture by Professor Gell-Mann in which he showed a slide of himself lecturing to an audience that included Dirac, fast asleep in the front row. So I hope you will notice that in this picture Kolmogorov is still awake—if slightly worried.

Also in 1922 Kolmogorov constructed the first example of an integrable function whose Fourier series diverges almost everywhere. He was only 19 years old at the time, and suddenly he had become an international celebrity—the more so after he

David Kendall is Emeritus Professor of Mathematical Statistics and Fellow of Churchill College, Department of Pure Mathematics & Mathematical Statistics, Statistical Laboratory, 16 Mill Lane, Cambridge BC2 1SB, United Kingdom.

FIG. 1. *Kolmogorov at the Tbilisi meeting, 1963.*

sharpened that result from almost everywhere to *everywhere*.

Three years later Kolmogorov wrote his first paper on probability, jointly with Khinchin. It contained a proof of the "three series" theorem, and the Kolmogorov inequality involving the maxima of partial sums of independent random variables (whence, ultimately, the martingale inequalities and the whole of the stochastic calculus).

Kolmogorov then became a doctoral student supervised by Luzin, and he emerged from this period of training with 18 mathematical papers to his credit. These contained the strong law of large numbers, the law of the iterated logarithm, some generalizations of the operations of differentiation and integration, and a contribution to intuitionistic logic. I am told that his two papers on this last topic are still regarded with awe by specialists in the field.

Four years later he began his lifelong friendship with Aleksandrov, marked by an expedition from Yaroslavl by boat down the river Volga and then on by way of Samara and the Caucasus (Figure 2) to Lake Sevan in Armenia. On the shores of the lake, Aleksandrov worked at his joint book on topology with Hopf, while Kolmogorov brooded over what was to be his 1931 paper on "Markov Processes with Continuous States in Continuous Time." Modern diffusion theory dates from that work, although it is analytical, and sample paths do not appear in it.

What was startling about the diffusion paper at the time of its appearance was the link with the theory of linear partial differential equations. Today, of course, the theory of parabolic and elliptic linear partial differential equations has merged with the theory of Markov processes, with each discipline lending strength to the other.

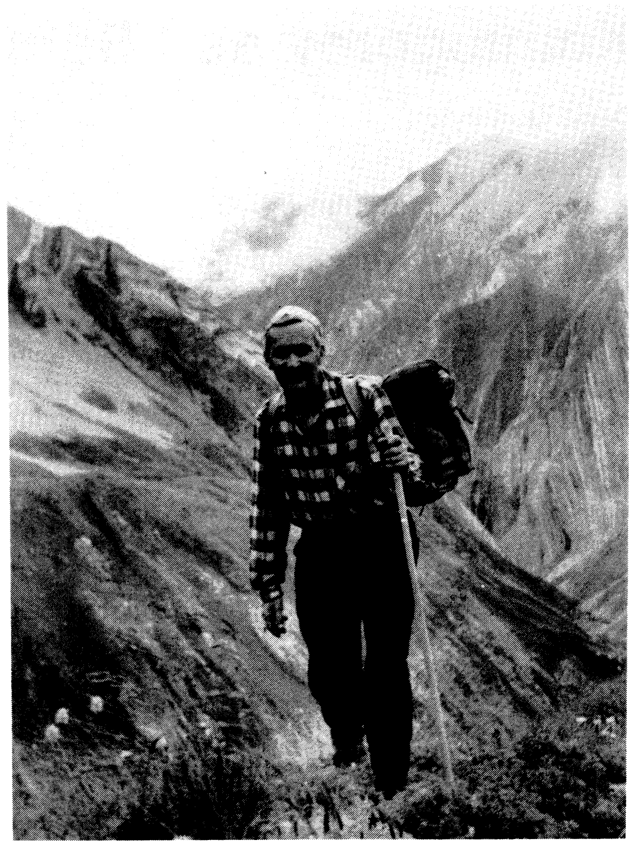


FIG. 2. *Kolmogorov in the Caucasus, about 1929.*

A little before this Kolmogorov had published his first attempt 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, but there were no sigma-algebras, no conditional expectations and no stochastic processes.

These omissions were however filled by Kolmogorov's famous monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung*. This was written in the forest on the banks of a small river and first published in the German language. In the foreword he explained that his intention was to create an *axiomatic foundation* for probability theory, and he remarked that without Lebesgue measure and integration this task would have been hopeless. He stressed that it is necessary first to strip away from the Lebesgue theory all those elements that tie it too closely to euclidean geometry, and he acknowledged the role that Fréchet played in bringing that about.

He directed the reader to three 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 explicit by the Romanian mathematician Ionescu Tulcea. Kolmogorov was careful to stress that the vital tool in the theory of conditioning is the generalization by Nikodym of an earlier theorem proved in a more classical setting by Radon. Another early treatment of infinite dimensional distribution theory is to be found in the papers of P. J. Daniell, but these seem to have attracted little notice until much later. However the historically important point is that *the proper development of stochastic process theory had to wait for a general treatment of conditioning*, and this Kolmogorov was the first to give. As he said, it would have been impossible without Nikodym's result.

In Kolmogorov's book the conditioning 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 his treatment of infinite-dimensional probability spaces, but it does not appear that the purely information-theoretic (i.e., nonprobabilistic) view of random events, random variables and σ-algebras had yet surfaced. This was to happen later, and it only became fully explicit in Doob's book on stochastic processes that appeared in 1953.

I like to recall a remark made by Kolmogorov during the International Congress of Mathematicians held in Amsterdam in 1954. A lunch for probabilists (held, perhaps appropriately, in the Zoo) had been organized 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 from now on be based on your work." I enjoyed watching Doob's pleasure, unsuccessfully concealed by embarrassment.

Some other anecdotes concerning Kolmogorov's respect for other mathematicians can suitably be related here. I have already mentioned his 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 always Kolmogorov—indeed we once spent a vacation together on the Mediterranean coast." Years later I referred to this when talking to Kolmogorov, and he said, "Pas exactement *un collègue*, plutôt *mon maître*."

On another occasion, in 1967, Hermann Dinges and I organized a meeting at Oberwolfach on the analytical theory of branching processes. We invited Kolmogorov, and to our delight he accepted and brought several other Soviet mathematicians with him. At first he explained that he only wanted to be a listener, but at the end of several highly mathematical talks he looked rather uncomfortable, and eventually told us that he would after all give a talk that would perhaps remind people of the biological background to the subject. Inevitably he referred to *The Genetical Theory of Natural Selection*—"das wundervolle Buch von R. A. Fisher." Two U.S. mathematicians whose identity I will not reveal were sitting near to me, and at this point I overheard one of them whisper to the other, "Well, it *can't* be the R. A. Fisher *we* know."

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

The "backwards" and "forwards" partial differential equations in the 1931 paper can be thought of as differentiated versions of what is now called the Chapman-Kolmogorov equation encapsulating the semigroup property characteristic of all markovian situations. I once asked Sydney Chapman about the origin of the phrase "Chapman-Kolmogorov" and was surprised to be told that he did not know of this terminology. Of course physicists have their own names for such things. The original Chapman reference seems to be his 1928 paper in the Royal Society's *Proceedings* concerned with the thermal diffusion of grains suspended in a nonuniform fluid.

It was in 1931 that Kolmogorov 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 talks with Lévy, while (as already mentioned) a long time by the sea was spent with Fréchet. Many years later Fréchet told me an amusing anecdote. They stayed in a lodging house whose proprietor had recently installed modern plumbing—unique in that small township. Thus the two mathematicians found themselves invited to a splendid party for the whole community, to celebrate—with champagne—the first flush.

In the 1930s Kolmogorov's work began 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 relating to these were being studied by a now growing school of colleagues and pupils—but it also saw Kolmogorov's independent development of cohomology theory, his theory of the structure and limiting behaviour of homogeneous countable Markov chains, his theory of statistical reversibility, his introduction of the characteristic functional (with a view to applications in nonlinear quantum mechanics), his inequalities for the absolute suprema of high derivatives (linked to the theory of quasi-analytic functions), his work with Gel'fand on rings of continuous functions on topological spaces and much else.

The "much else" included contributions to queueing theory, to branching process, to the stochastic geometry of the crystallization process (and of the growth of vegetation), 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 community. This last was written within a year of a similar but independent study by Fisher. Both papers reported that the range inhabited by the genetically favored individuals would expand with an asymptotically constant velocity, but Kolmogorov and his two colleagues showed that in fact there is a half-infinite interval of possible speeds with each of which there is associated a corresponding traveling wave. This is now a subject in its own right called *reaction-diffusion theory*. The application to genetics with which the subject originated has since been joined by applications to the spreading of epidemics, to the spreading of cultural innovations, to the dynamics of advertising, to the spreading of rumors and to numerous other physical, chemical and biological problems.

Kolmogorov's immensely influential work on the *smoothing and prediction* of stochastic processes with stationary ordinates (or increments) started as early as 1938 with a paper written against a background provided by Khinchin and Slutskii. These topics proved later to be of very great military importance, and so it is scarcely surprising that another attack on the problem was mounted by Norbert Wiener in the United States. Such investigations were eventually to be covered by a cloak of secrecy during World War II, and only a privileged few were allowed to know about them. No one in the West seemed to be aware that the basic ideas had already been openly published by Kolmogorov.

It can now be seen that the two approaches followed respectively by Kolmogorov and Wiener complemented one another in an interesting way, and this has been analyzed by Peter Whittle in a recent survey. Insofar as priority in such a confused situation is important, there seems to be no doubt that it was Kolmogorov who was first in the field.

This work eventually brought about a profound change in the relationship 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 that now affects almost every aspect of our lives.

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

From stationary stochastic processes to stationary stochastic fields and thence to the study of *turbulence* is a natural progression. Kolmogorov's interest in turbulence dated from the late thirties, and it was to lead to one of his greatest discoveries. In 1940 he wrote a famous paper on the local structure of turbulence, and this was later supplemented by what is now called 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, and in 1970-1972 he sailed around the world with the scientific research ship, the *Dmitrii Mendeleev*, as Scientific Supervisor of a study of oceanic turbulence (Figures 3 and 4). This allows us to think of Kolmogorov as in some sense a colleague of Edmund Halley and James Cook.

In the immediate post-war period we find Kolmogorov writing on mathematical geology, inferential statistics and branching processes, and at the same time contributing 88 articles to the Soviet encyclopedia, as well as working with Gnedenko on their book *The Limit Distributions for the Sums of Independent Random Variables*, which was to become another 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 one on continuous-time Markov chains written in 1951 was quite different. That revealed by way of examples certain bizarre phenomena (originally called "pathological") and asked for their investigation. This paper appeared at about the same time as a complementary one by Lévy, and the two papers together have generated a huge literature. Harry Reuter and I immediately set about trying to understand Kolmogorov's examples, spurred on in this task by Kai-lai Chung. This was for us the beginning of a very enjoyable collaboration that is something we personally owe to Kolmogorov.

This work led eventually to the following unsolved problem. Let $p(t)$ and $p^*(t)$ be two matrices of standard Markov transition functions, and suppose it is known that $p_{ij}(t) = p^*_{ij}(t)$ throughout a nondegenerate time interval $[0, T_{ij}]$ for all values of i and j. *Does it then follow that the two matrices of Markov transition functions are identical?*

Substantial progress has been made with this question (confirming the truth of the conjecture in a number of special cases) by Di San-min, Hou Zhen-ting and Yu Yao-qi, and the implications of their work have been summarized in a recent review by Reuter, but we still do not know the answer in general.

More recently Sofia Kalpazidou has launched a spirited attack on the problem within the context of "cycle theory." Yet another possibility might be to reformulate the problem in the language of nonstandard analysis. Of course one's hope is that, if the conjecture is generally true, then one might be able to reformulate the infinitesimal generator of the process in terms of the "germs" of the functions p_{ij} at $t = 0$.

2. KOLMOGOROV AND THE HILBERT PROBLEMS

One of the most famous of Kolmogorov's achievements outside the field of probability theory was his proof that every continuous real function $f(x_1, x_2, \ldots, x_n)$ can be expressed in the "slide-rule" form

$$\sum_{j=1}^{2n+1} g_j\left(\sum_{i=1}^{n} \phi_{i,j}(x_i)\right),$$

where the functions g_j and $\phi_{i,j}$ are continuous *and only the g-functions depend on f*. However my reference to slide-rules is not really legitimate, because for that to be appropriate we would need all the ϕ's and g's to be (at least piecewise) smooth and monotone.

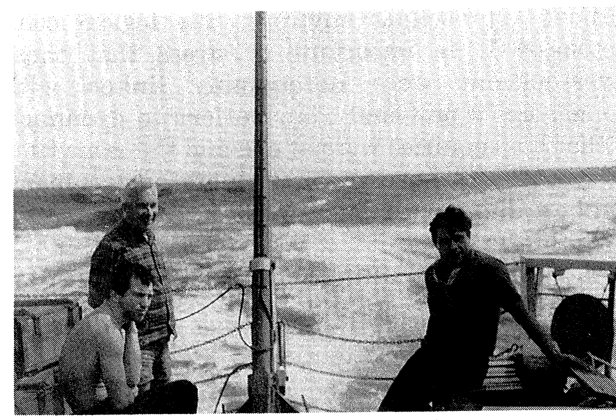


FIG. 4. *Kolmogorov on the Dmitrii Mendeleev in the Atlantic, west of Gibraltar, 1971.*

FIG. 3. *Kolmogorov on the Dmitrii Mendeleev southwest of the English Channel, 1971.*

This result solved one of the famous Hilbert problems, as did also Kolmogorov's definitive formulation of probability theory in 1933. Indeed, as the reader will see below, there is a sense in which Kolmogorov solved the latter problem *twice*!

3. FROM PROBABILITY TO COMPLEXITY

From the 1950s onwards Kolmogorov's most important scientific work revolved around the quartet of ideas: *probability*, *dynamics*, *information* and *complexity*. Of all his work this is perhaps the most difficult and the most important. I will therefore try to sketch some of its more interesting features, without approaching anything like logical completeness. It is important to stress that these investigations were indissolubly linked with Kolmogorov's profound contributions to dynamics, so that his enquiries were at one and the same time concerned with how we perceive our environment, and how that environment works.

4. A NEW ENVIRONMENT FOR PROBABILITY

When one re-reads Kolmogorov's *Grundbegriffe* of 1933, with proper attention to the footnotes, one is much struck by what he does *not* say. (Those familiar with the memoirs of Mr. Sherlock Holmes will at once recall the curious incident of the dog in the night-time.) Kolmogorov makes many interesting comments, but he seems in that book to shy away from any detailed explanation of the relationship between his axioms and empirical practice, referring the reader to von Mises for that.

A careful study of the chronology suggests that his questing mind was already wrestling with the deep philosophical problems associated with the notion of randomness, and I now believe that this must have been one of the main topics of his long sea-side discussions with Fréchet, especially as the books by Hostinský and von Mises had just appeared. In particular Hostinksý's revival of Poincaré's "explanation" of randomness in terms of the discrete (if fantastically fine) partitioning of dynamical phase space must have been a source of much of Kolmogorov's later thinking.

In the end these influences led to a re-formulation of probability theory and information theory that is almost a cultural revolution, *turning each of these subjects inside out, and reversing the order in which they are normally considered.*

There are philosophical (and indeed practical) aspects of this work with which we must all become familiar, because it is clear that the new point of view is likely to percolate throughout the whole of science. For a detailed presentation I have to refer you to his own all too few written accounts and to the very important related work by others. My own account is merely one written by an onlooker, but possibly such an informal review may prove to be useful, at least to some other onlookers.

The first thing to realize is that the theory is based on the consideration of *finite* objects and *finite* algorithmic operations on them. Its spirit is summed up in a quotation from Kolmogorov's 1963 article 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 that 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 Kolmogorov 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 necessary to add another of Kolmogorov's remarks:

> 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.

So now let us try to catch at least the gist of this new approach, viewed here for the sake of simplicity in a typical "context," that of Lebesgue measure on the Borel subsets of the set of all $(0, 1)$-sequences. (In the language of the *Grundbegriffe* we could equivalently say that we have in mind an infinite sequence of Bernoulli trials with individual chance $p = \frac{1}{2}$. 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 specific triple $(\Omega, \mathscr{F}, \mu)$ with $\mu(\Omega) = 1$ in the usual way, and in what follows reference to the model will be indicated by a mention of "the context.")

I will rely heavily on a survey by Kolmogorov and Uspensky that was presented by the latter at the First World Congress of the Bernoulli Society, and on a subsequent paper by Vovk, and I will quote freely from each of these papers.

The first step is to introduce four sets of infinite $(0, 1)$-sequences $\omega = (\omega_1, \omega_2, \ldots)$, to be called **T**, **C**, **KS** and **CS**. The "definitions" given here are informal only, and omit essential detail far beyond our present scope.

A given infinite $(0, 1)$-sequence $\omega = (\omega_0, \omega_1, \ldots)$ will be a member of **T** (and then called *typical*) if and only if it belongs to every set of $(0, 1)$ sequences that effectively has measure 1. It is a theorem of Martin-Löf that **T** defined in this way effectively has *itself* measure 1, so that it is the smallest such subset. Here "effective" refers to the explicit algorithmic basis of the whole approach. So when ω belongs to **T** then "it belongs to every reasonable majority."

A given sequence ω will be a member of **C** (and then called *chaotic*) 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*. So membership of **C** means that there is no simpler way of describing the alternation of 0's and 1's.

This last definition assumes that we have already defined Kolmogorov's so-called "optimal monotone complexity" $K(\theta)$ for each fixed *finite* object θ, as always relative to the context. Here again we omit the details. The basic idea is that the *complexity is essentially the length of the shortest possible description*.

To elaborate a little, I remark that we can suppose the length of the description of the finite object $(\omega_1, \omega_2, \ldots, \omega_n)$ to be no greater than n, and "growing at the fastest possible rate" can be taken to mean that $K(\omega^n)$ never falls short of n by less than some $c(\omega)$ *that is independent of* n. A fundamental theorem guarantees that if we choose a different optimal monotone complexity then the result is not essentially altered.

A theorem of Levin and Schnorr now tells us 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 (2) *nontypical and nonchaotic*.

Accordingly what we shall call **R** (for "random") $= \mathbf{T} = \mathbf{C}$ can be taken to be *the natural home of those* 0-1 *sequences that form the basis of probability theory*. To justify this, however, we must also bring in two other sets **CS** and **KS** of infinite $(0, 1)$-sequences. (Here **S** stands for "stochastic.") These are formulated in language similar to that used by von Mises when describing his "collectives."

The first version, **CS** (**C** for Church), consists of those 0-1 sequences for which the frequency-ratio converges to p (here $\frac{1}{2}$) in every *effectively* selected subsequence.

The set **KS** (**K** for Kolmogorov) is defined similarly, but here we now require convergence to p in all the subsequences that can be built up when at each stage one is allowed to select *any* symbol in the sequence that has not already been chosen—i.e., one is allowed (effectively) to "dodge about" when selecting new terms.

There is then a generalization of the Levin–Schnorr theorem asserting that

$$\mathbf{T} = \mathbf{C} \subset \mathbf{KS} \subset \mathbf{CS},$$

and work by Loveland tells us that the last inclusion is strict, so that **CS** can be ignored in what follows.

Accordingly we can use $\mathbf{C}(= \mathbf{T} = \mathbf{R})$ to provide an environment in which to do classical probability with a new—an *entropic*—motivation.

Of course it is natural to ask if the first inclusion, $\mathbf{C} \subset \mathbf{KS}$, is an equality. In one of his publications Kolmogorov announced that **KS** is strictly larger than **C**, but his proof of that assertion has been lost. However a letter from Razborov tells me that a proof has now been supplied by Shen, so we know that **KS** really is strictly larger than **C**.

There are still some unsolved problems relating to **KS**. Leaving these on one side, we can regard the statement $\mathbf{C} \subset \mathbf{KS}$ as a theorem assuring us that a strong form of the von Mises "stochastic" property holds in **C**.

Thus the class **C** of chaotic 0-1 sequences becomes a natural environment in which to do probability calculations.

Perhaps we will have to think of **KS** as a technical enlargement of **C** facilitating certain calculations, just as the class of Lebesgue sets affords a technical enlargement of the more comprehensible class of Borel sets.

To practical probabilists many other questions 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 that done in the near future.

One can now proceed to rebuild probability theory starting with the infinite $(0,1)$-sequences in **C**, and so *using entropic instead of probabilistic methods*. Kolmogorov was convinced that this was possible, and indeed he knew that to some extent it had been carried out. It is therefore especially fitting that when the last of Kolmogorov's papers (written jointly with Uspensky) was published in *Teoriya*, it was immediately followed by a remarkable paper by Vovk that exemplifies in a triumphant manner the success of this part of the Kolmogorov programme.

5. VOVK'S THEOREMS

Vovk gives us nothing less than an entropic proof of the classical law of the iterated logarithm for an arbitrary *fixed* and *suitably nearly chaotic* infinite $(0,1)$-sequence ω. As that theorem first emerged in the context of number theory, this is natural enough, but one wonders what G. H. Hardy would have thought of it.

Figure 5 is a diagram that may elucidate the situation. It indicates the locations of a number of famous theorems on the complexity scale.

The convention is that complexity increases from the bottom to the top of the picture. At the top we have **C**, the regime of maximal entropy and complexity. This we can think of as a home for the iid $(0,1)$-sequences of classical probability theory.

At the bottom of the picture we have the "boring" sequences of zero entropy and zero complexity.

In between we have sequences in which $K(\omega^n)$ lags behind the maximum value n by some function $f(n,\omega)$ that is of smaller order than n itself.

—Maximum complexity—

- -

LIL : $f(n,\omega) = o(\log\log n)$

REC : $f(n,\omega) = (\frac{1}{2} - \epsilon)o(\log_2 n)$

SLLN : $f(n,\omega) = o(n)$

- -

—Minimum complexity—

FIG. 5. *The scale of complexity.*

In the picture we have three such levels, the first (a very high one) with

$$f(n,\omega) = o(\log\log n),$$

the second (an intermediate one) with

$$f(n,\omega) = (\tfrac{1}{2} - \epsilon)o(\log_2 n) \ (\epsilon > 0)$$

and the third (a very low one) with

$$f(n,\omega) = o(n).$$

Vovk has shown that the strong law of large numbers holds *at and above the lower of these three levels*, that the recurrence property for equal numbers of 0's and 1's holds *at and above the intermediate level* (for every choice of ϵ) and that the law of the iterated logarithm holds *at and above the highest of the three levels*.

Volk's paper contains much more information than this, but what I have extracted from it will suffice for our present purposes—and is indeed as far as my own knowledge goes. Now doubt the subject will develop rapidly after this fine start, and what is reported here may well already be out of date.

To sum up: We can now assert analogs of the classical probability limit theorems for *suitably nearly chaotic* individual infinite $(0,1)$-sequences, and also we can begin to classify such theorems according to *the degree of complexity 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. G. H. Hardy would indeed have been interested!

For the three examples located in the picture the degrees of complexity are simply ordered, and it is natural to ask whether this will always be so. I do not know the answer to that question, but expect it to be "No."

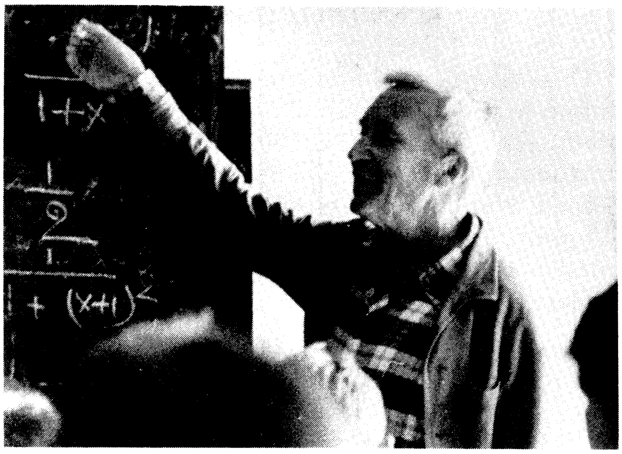


FIG. 6. *Kolmogorov teaching in his school, 1968.*

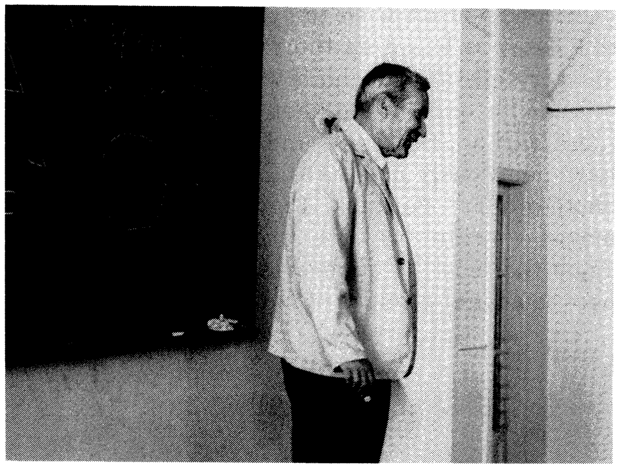


FIG. 7. *Kolmogorov talking to school children.*

6. KOLMOGOROV AS A TEACHER

At all stages of his career Kolmogorov 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 occupied with his growing interest in mathematical education, taking very heavy responsibilities in connection with one of the special schools for gifted children sponsored by the Moscow State University. To this school he devoted a major portion 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. Kolmogorov (Figure 6) sought for these children a broad and natural development of the whole personality. It did not worry him if they did not become mathematicians. Whatever profession they ultimately followed, he was content if their outlook remained broad and their curiosity unstifled. Indeed it must have been wonderful to belong to this extended family of Andrei Nikolaevitch. See him talking to some members of it (Figure 7).

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

7. CONCLUSION

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

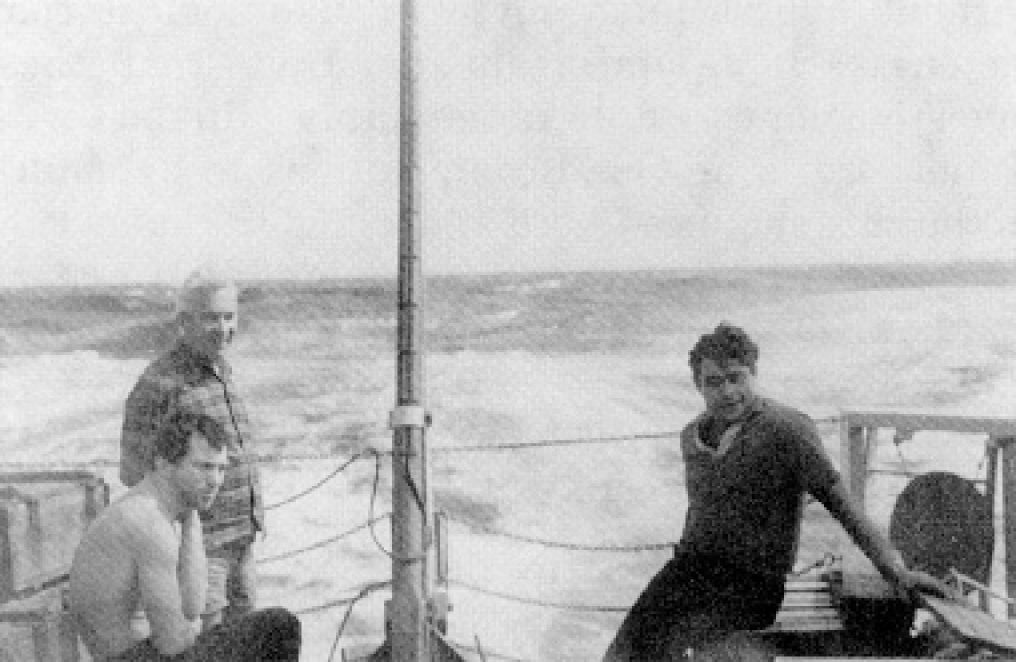
FIG. 8. *Kolmogorov and the seal (Galapagos Islands, 1971).*

FIG. 9. *Kolmogorov in the Crimea (Simferopol, 1970).*

In conclusion I wish to acknowlege how much of this essay is based on the recollections of others. I am very grateful to those who have given me permission to quote from their writings, and also to those (Albert Shiryaev and Igor Zhurbenko) who have generously made available to me the splendid photographs. Perhaps Figure 8, showing Kolmogorov swimming with a seal, is the one you will enjoy the most. (Kolmogorov is on the right.) And here is a final glimpse of Kolmogorov (Figure 9).

ACKNOWLEDGMENTS

This essay is an abbreviated version of my contribution to the full-length obituary in *Bull. London Math. Soc.* 22 (1990) 31-100, and it coincides approximately with my Kolmogorov Lecture at the Second World Congress of the Bernoulli Society (Uppsala, 1990). It is published here with the permission of the London Mathematical Society, the holder of the copyright.




1+x
(x+1)