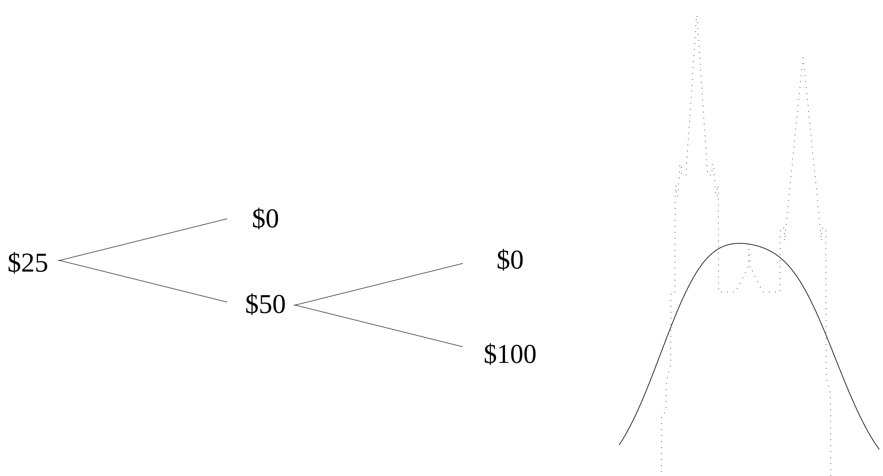
Kolmogorov's contributions to the foundations of probability

Vladimir Vovk
and Glenn Shafer



The Game-Theoretic Probability and Finance Project

Working Paper #6

January 27, 2003

Project web site:
http://www.probabilityandfinance.com

Abstract

Andrei Nikolaevich Kolmogorov was the foremost contributor to the mathematical and philosophical foundations of probability in the twentieth century, and his thinking on the topic is still potent today. In this article we first review the three stages of Kolmogorov's work on the foundations of probability: (1) his formulation of measure-theoretic probability, 1933, (2) his frequentist theory of probability, 1963, and (3) his algorithmic theory of randomness, 1965–1987. We also discuss another approach to the foundations of probability, based on martingales, that Kolmogorov did not consider.

Contents

1 Introduction

The exposition of this paper is based on Figure 1. At the center of the figure is Kolmogorov's earliest formalization of the intuitive notion of probability, in his famous *Grundbegriffe der Wahrscheinlichkeitsrechnung* in 1933. This formalization, *measure-theoretic probability*, has served and still serves as the standard foundation of probability theory; virtually all current mathematical work on probability uses the measure-theoretic approach. To connect measure-theoretic probability with empirical reality, Kolmogorov used two principles, which he labeled A and B. Principle A is a version of von Mises's requirement that probabilities should be observed frequencies. Principle B is a finitary version of "Cournot's principle", which goes back to Jacob Bernoulli's *Ars Conjectandi* (1713) and was popularized in the 18th century by Antoine Cournot.

The goal of Kolmogorov's later attempts to formalize probability was to provide a better mathematical foundation for applications. In §3 we discuss his frequentist theory of probability and in §4 his algorithmic theory of randomness.

Another strand of work also derived from von Mises's ideas. In his 1939 book *Étude critique de la notion de collectif* Jean Ville proposed an improvement on von Mises's approach that used game-theoretic ideas going back to his famous compatriot Blaise Pascal. The resulting notion of martingale was never used by Kolmogorov in his studies of the foundations of probability; he developed von Mises's definition in directions quite different from Ville's. In §5 we discuss this strand of research, including our recent suggestion [29] to base the mathematical theory and interpretation of probability directly on the notion of martingale.

We conclude by discussing the usefulness of Kolmogorov's algorithmic randomness. We argue that, even though its usefulness as a framework for stating new results about probability might be limited, it has a great potential as a tool for discovering new facts. We describe in detail an example from our own research.

We cover both mathematical and scientific aspects of Kolmogorov's work. Since the mathematical aspects are uncontroversial and generally well-known, our emphasis is often on the scientific aspects (how mathematical theories developed by Kolmogorov relate to reality). But the division of our discussion into sections is according to the mathematical theories of probability.

2 Measure-theoretic probability

According to *Grundbegriffe*, the mathematical theory of probability studies probability measures, i.e., measures P on a measurable space $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$. An event is just a set $E \in \mathcal{F}$ and its probability is $P(E)$. On this simple foundation Kolmogorov built a rich mathematical theory which has been developed by many researchers and is well known. In this short section we are mainly interested in connections of measure-theoretic probability with reality, as Kolmogorov saw them.

Kolmogorov offered two principles for interpreting the probability $P(E)$,

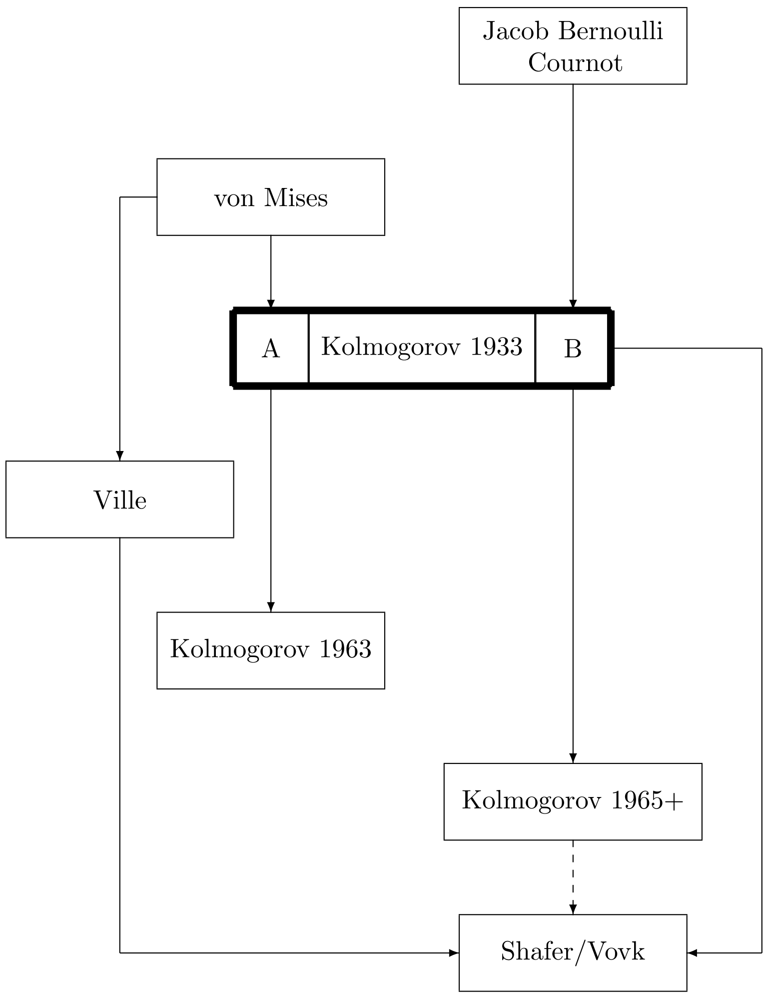


Figure 1: Some attempts at formalization and interpretation of probability (in chronological order from top to bottom)

where E is an event that happens or does not happen in experiment C:

A. One can be practically certain that if C is repeated a large number of times, the relative frequency of E will differ very slightly from $P(E)$.

B. If $P(E)$ is very small, one can be practically certain that when C is carried out only once, the event E will not occur at all.

Each of these two principles has a rich history. In A, Kolmogorov follows the frequentist ideas of Richard von Mises (specifically referring to his [24]). Principle B goes back to Bernoulli, Cournot, and Lévy (see [29]).

3 Frequentist probability

Kolmogorov published several informal expositions of his frequentist philosophy of probability in 1938–1959 [11, 12, 13, 14, 20]. His only attempt to formalize this philosophy came in 1963, in *On tables of random numbers* [15]. He starts this article by stating his reasons for not making such an attempt earlier:

1. The infinitary frequency approach based on *limiting frequency* (as the number of trial goes to infinity) cannot tell us anything about real applications, where we always deal with finitely many trials.

2. The frequentist approach in the case of a large but finite number of trials cannot be developed purely mathematically.

Kolmogorov never changed his mind about point 1. But by 1963 his thinking about the complexity of algorithms had led him to change his mind about point 2. He now saw that he could use the fact that there are few simple algorithms to define a finitary version of von Mises's collectives.

Consider a finite sequence $(x_1, \ldots, x_N)$ consisting of ones and zeroes. What properties should this sequence have in order for us to regard it as random (intuitively, the result of independent trials)? Von Mises had tried to answer this question for an infinite sequence; he had said that an infinite sequence is random (or that it is a "collective") if it satisfies two requirements:

1. The limiting frequency of ones exists.

2. This limiting frequency does not change if we choose an infinite subsequence without knowing the outcomes in advance.

In the case of finite sequences the first condition is vacuous, and we only need to worry about the second one. Of course the words "does not change" must be replaced with "does not change much", which leads to the need to speak about (N, ϵ)-random sequences rather than just random sequences. In fact, Kolmogorov's definition in [15] is even more complicated, since there is an extra parameter $\mathcal{R}_N$ (the set of admissible selection rules, themselves complicated objects), which is not explicitly reflected in his notation. This lack of mathematical elegance was probably balanced, for Kolmogorov, by the philosophical importance of the frequentist concept.

Kolmogorov mentions in [15] that he is interested in *simple* selection rules. He does not define simplicity precisely, but he relies on the assumption that there are not too many selection rules in $\mathcal{R}_N$, and he explains that this is justified by the fact that there cannot be many simple selection rules.

Although his definition of (N, ϵ)-randomness is clearly intended to serve as the basis for a finitary version of von Mises's theory, Kolmogorov did not go on to develop such a theory. In fact, he dropped the concept altogether in favor of a more direct and elegant approach to defining randomness, to which we now turn.

4 Finitary algorithmic randomness

As we have just seen, Kolmogorov's was thinking in 1963 about using algorithmic complexity as the starting point for a frequentist definition of probability, in the style of von Mises. By 1965, however, he was thinking about defining randomness directly in terms of algorithmic complexity, in a way that also allows us to connect randomness more directly to applications, without any detour through the idea of frequency. We call this approach Kolmogorov's *finitary theory of algorithmic randomness*. He first mentioned it in print in 1965, in the final paragraph of *Three approaches to the definition of the concept "amount of information"* [16]. He developed it further in articles published in 1968 [17] and 1983 [18, 19]. His most detailed exposition was in *Combinatorial basis of information theory and probability theory* [18], published in 1983 but first prepared in 1970 in connection with his talk at the International Mathematical Congress in Nice.

Bernoulli sequences

Suppose the binary sequence $(x_1, \ldots, x_N)$ has k ones and $N - k$ zeroes. To describe this sequence, $\log \binom{N}{k}$ bits are sufficient, since there at most $\binom{N}{k}$ such sequences. According to Kolmogorov, the sequence is *Bernoulli* if it cannot be described with substantially fewer bits.

To make this definition precise we need to define the shortest description of a sequence. The key discovery that made this possible was that there is a *universal* method of description, which provides descriptions almost as short as those provided by any alternative method. (This was realized independently and earlier by Ray Solomonoff [31, 32]. The existence of a universal method of description is an implication of the existence of a universal algorithm.)

The *Kolmogorov complexity* $K(x)$ of x is defined to be the length of the shortest description of x when the universal method of description is used. Allowing the universal method of description to use extra information y, we obtain the definition of conditional Kolmogorov complexity $K(x \mid y)$. Now we can say that a sequence $x = (x_1, \ldots, x_N)$ with k ones is *Bernoulli* if $K(x \mid N, k)$ is close to $\log \binom{N}{k}$. Kolmogorov made this precise by giving a real number m that measures the closeness: the sequence is *m-Bernoulli* if it satisfies

$$K(x \mid N, k) \geq \log \binom{N}{k} - m.$$

Interpretative assumption

A Bernoulli sequence is only one example of a random object. The general concept, already present in the brief comment in the 1965 article, is that of a random object in a large finite set A, which must itself be given by some finite description. Given the description of A, the complexity of an element is at most the logarithm of the number of elements of A. A *random* element x is one whose complexity is close to this maximum—i.e., one satisfying

$$K(x \mid A) \approx \log |A|. \tag{1}$$

Formally, we can call the difference $\log |A| - K(x \mid A)$ the *deficiency of randomness* of x in A.

Recall that Kolmogorov's measure-theoretic probability was connected to the empirical world by two interpretative assumptions, Principles A and B. The main interpretative assumption of his finitary theory of algorithmic randomness is that we expect the realized outcome x of a trial with outcomes in a finite set A to be random in A in the sense of (1).

Statistics with complexity models

The basis of standard statistics is the notion of a *statistical model*, i.e., a family of probability distributions for the outcomes of an experiment. Kolmogorov suggested that we replace statistical models with what we call in this article *complexity models*: classes of disjoint sets whose union contains all possible outcomes of the experiment. (The experiment may be very complex. Typically it consists of a sequence of trials of a more elementary experiment.)

We apply Kolmogorov's interpretive assumption to a complexity model by assuming that the outcome of the experiment will be random with respect to the set in the model that contains it. (Because the sets in the model are disjoint, there is exactly one such set.) In other words, one's acceptance of a complexity model is interpreted as one's belief that the randomness deficiency of the actual outcome in the set of the model to which it belongs will be small.

Kolmogorov has given several examples of complexity models, which we reproduce here. All sequences in our descriptions are assumed to be finite.

Example 1 A binary sequence x is Bernoulli if

$$\left.\begin{aligned} N &= \text{length of } x \\ k_0 &= \#\text{ of 0s in } x \\ k_1 &= \#\text{ of 1s in } x \end{aligned}\right\} x \text{ is random given } N, k_0, k_1.$$

As we have already explained, this means that the complexity $K(x \mid N, k_1)$ is close to $\binom{N}{k_1}$. Formally, the Bernoulli complexity model consists of all equivalence classes of finite sequences of ones and zeroes, where the equivalence of two sequences means that they have the same length and the same number of ones (and hence also the same number of zeroes).

Example 2 A binary sequence x is *Markov* if

$$\left.\begin{aligned} N &= \text{length of } x \\ s &= \text{1st element of } x \\ k_{00} &= \#\text{ of 00 in } x \\ k_{01} &= \#\text{ of 01 in } x \\ k_{10} &= \#\text{ of 10 in } x \\ k_{11} &= \#\text{ of 11 in } x \end{aligned}\right\} x \text{ is random given } N, s, k_{00}, k_{01}, k_{10}, k_{11}.$$

In analogy with the previous example, the Markov complexity model consists of all equivalence classes, where the equivalence of two sequences means that they have the same length, the same first bit and the same number of transitions $i \to j$ for all $i, j \in \{0, 1\}$.

Example 3 Markov sequences x of order d: defined analogously to the previous example.

Example 4 A sequence $x = (x_1, \ldots, x_N)$ of real numbers is *Gaussian* if

$$\left.\begin{aligned} N &= \text{length of } x \\ m &= \text{sample mean } \tfrac{1}{N}\textstyle\sum_{n=1}^{N} x_n \\ \sigma^2 &= \text{sample variance } \tfrac{1}{N}\textstyle\sum_{n=1}^{N}(x_n - m)^2 \end{aligned}\right\} x \text{ is random given } N, m, \sigma^2.$$

To make this precise, one must, of course, discretize the real line in some way.

Example 5 In a similar way one can define *Poisson* sequences.

Some mathematical results about complexity models are proven in Eugene Asarin's articles [1, 2] and his PhD thesis [3] (supervised by Kolmogorov). A typical result is of the following form: if x is Gaussian (see Example 4), N large and $[a, b]$ a fixed interval,

$$\frac{\#\{x_n \in [a, b]\}}{N} \approx \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt.$$

The idea of a complexity model eliminates probability as the basic notion for the foundations of statistics. We can reintroduce probability by calling certain relative frequencies probabilities. In Example 2, for instance, we can define the conditional probability that a 0 will be followed by a 1 in the Markov sequence x to be the ratio of the number of occurrences of 01 in x to the total number of occurrences of 00 and 01 in x. This step is very natural in view of Kolmogorov's original motivation, but we can use complexity models directly for prediction without talking about probability. Suppose, e.g., that we are willing to make the assumption that a sequence of real numbers of length N, only half of which is known yet, is Gaussian. Then we can predict that the mean value of the numbers in the second half will be close to that in the first half. It is clear that many other predictions of this type can be made for the Gaussian model and the other models we have considered. No probabilities are needed in these applications.

Stochastic sequences

If $(x_1, \ldots, x_N)$ is Bernoulli, then it is random in a simple set (the set of all sequences with the same length and same number of ones can be described using approximately $2 \log N$ bits of information). This is true of the other models we have listed as well, and this seems to be the source of the predictive power of these models. At a 1982 seminar in Moscow University Kolmogorov defined a finite object x to be (α, β)-*stochastic* (where α and β are, typically small, positive integers) if there exists a finite set A such that

$$x \in A, \quad K(A) \leq \alpha, \quad K(x \mid A) \geq \log |A| - \beta.$$

The first to study Kolmogorov's notion of stochasticity were Shen' [30] and V'yugin [37]; they answered the question of how many (in different senses) stochastic sequences are there. There are many other interesting mathematical questions, both answered and open, raised by Kolmogorov's concept of stochasticity; see, e.g., V'yugin [38], Li and Vitányi [22], Subsection 2.2.2, Gács et al. [9], and Vereshchagin and Vitányi [33].

Is the new theory frequentist?

Of the complexity models we have listed, only Example 1 is of direct interest for the frequentist interpretation of probability. The other examples are natural developments, but it is clear that they go beyond frequentism. Moreover, as we have already remarked, the concepts of probability and frequency are not needed when we use complexity models for prediction.

Kolmogorov never renounced frequentism, but it is clear that the new theory is not frequentist in the sense of von Mises. Its essential characteristics are the following:

1. It is strictly finitary: only finite sequences and finite sets of constructive objects are considered.

2. It makes an assumption analogous to Kolmogorov's Principle B, which said that an event of very small probability will not happen. Now we say

that a particular event, the event that the realized outcome is very simple, will not happen.

Principle A, the principle of frequency, does not play an essential role. No do probability distributions play any role; they are replaced by finite sets.

Martin-Löf's development

In 1966 Martin-Löf [23] gave another justification of Kolmogorov's notion of randomness deficiency, showing that it is a universal statistical test. The notion of a universal statistical test was easy to extend to the case of infinite sequences. It was later shown, by Levin [21] and Schnorr [28], that a sequence is Martin-Löf random if and only if the randomness deficiency of its initial fragments is bounded (the randomness deficiency, however, must be defined in terms of some variant of Kolmogorov complexity, such as "monotonic" or "prefix" complexity).

5 Martingales

Kolmogorov's frequentist approach as developed in [15] was based on von Mises's ideas. In his 1939 book Jean Ville found, however, that von Mises's definition of collective, as formalized by Wald [39], has a serious shortcoming: there are collectives that violate the law of the iterated logarithm; this is also true about Church's 1940 formalization. Kolmogorov was interested in finite sequences, but Ville's example has unpleasant implications for long finite sequences as well. In [15] Kolmogorov solves this difficulty by allowing subsequence selection rules to scan the sequence in arbitrary order, but it turned out that the solution offered by Ville himself was a groundbreaking discovery originating a new and fruitful area of probability theory. The justification offered by von Mises for his notion of collective was the "principle of the excluded gambling system"; Ville generalized in a very natural way von Mises's notion of a gambling system, introducing the notion of martingale and modifying the notion of collective. The crucial fact proved by Ville was that his example, or an analogous example based on any other strong limit theorem, was impossible for the modified collectives. (We already mentioned a similar property of universality established by Martin-Löf [23] for Kolmogorov's later definition based on Kolmogorov complexity.)

Towards 1971 Schnorr [27, 26] suggested several definitions of randomness through martingales. A similar theory, but without explicit use of martingales, was developed by Levin (cf. Levin's [21] definition of randomness in terms of *a priori* probability).

Doob [8] expressed and developed Ville's ideas inside measure-theoretic probability, and now martingales are central to that theory. The notion of martingale is, however, obviously more game-theoretic than measure-theoretic. In our book [29] we recount the history of game-theoretic ideas in the foundations of probability, going back to Pascal, and we show how the classical core of probability theory can be based directly on game-theoretic martingales, with no appeal to measure theory. Probability again becomes secondary concept but is now defined in terms of martingales; we show that it can serve well for many purposes. This approach is well suited not only to the classical limit theorems but also to many applications of probability to finance. In the next section we demonstrate the approach with an example.

6 The heuristic value of algorithmic randomness

Kolmogorov's algorithmic theory of randomness continues to generate very interesting new research (see, e.g., [9, 33]). However, the greatest value of the algorithmic theory may be as a tool of discovery rather than as a language for mathematical exposition or practical application. Because it is so precise from the intuitive point of view, the theory is of great use in discovery. But the details of the mathematical apparatus that makes the theory so precise and explicit—the additive constants, computability in the limit, etc.—get in the way as soon as we turn to exposition and application. So as soon as a result is proven, there is much to be gained by removing algorithmic ideas from that result. Algorithmic randomness thereby disappears, and its guiding role is hidden from the future users of the result.

Because the algorithmic results are usually not even published, it is not easy to give examples of this transformation within the confines of an expository article. For example, we can point to many of the results in our book [29] as examples of algorithmic results that were stripped down to a game-theoretic form, but few readers will find this enlightening in the absence of any published or even well-polished exposition of the original algorithmic results. We will, however, give one example. For simplicity, and leaving aside Kolmogorov's philosophical preferences, we make this example infinitary.

Infinitary algorithmic probability theory, as developed by Martin-Löf, permits a sharp distinction between random and non-random infinite sequences, leading to pointwise strong limit theorems. For example, we can restate Borel's strong law of large numbers by saying that all random infinite binary sequences $x_1 x_2 \dots$ satisfy

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{2}. \tag{2}$$

Schnorr [25, 27] quickly noticed that (2) also holds for many non-random sequences as well; actually, for (2) to hold it suffices to require a subexponential rate of growth of the randomness deficiency of the initial fragments of $x_1 x_2 \dots$. This observation was extended in [35] to two other strong limit theorems: the law of the iterated logarithm and the recurrence property. The requirement on the rate of growth of randomness deficiency is much stronger for these two finer laws: for example, one of the results of [35] was that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n \ln\ln n}} \left| \sum_{i=1}^{n} x_i - n/2 \right| = \frac{1}{\sqrt{2}}$$

if the rate of growth is $o(\ln\ln n)$; on the other hand, for any function $f(n) \to \infty$, $n \to \infty$, there exists a sequence with the rate of growth $f(n) \ln\ln n$ such that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n \ln\ln n}} \left| \sum_{i=1}^{n} x_i - n/2 \right| = \infty.$$

As Schnorr [26] explains, the fastest admissible rate of growth of randomness deficiency indicates the importance of the limit theorem under consideration; the fact that for the strong law of large numbers the rate can be almost as fast as exponential reflects the fact that this law is one of the most basic limit theorems of probability.

We will now explain how the algorithmic idea of classification of the strong limit theorems can be expressed in the game-theoretic framework of [29].[1] Instead of a binary sequence $x_1 x_2 \ldots$ we consider a bounded (by 1, without loss of generality) in absolute value sequence of real numbers and also allow a Forecaster to announce a predicted value m_n for each x_n. Formally, we consider the following protocol.

Bounded Forecasting Game
Players: Forecaster, Skeptic, Reality
Protocol:
$\mathcal{K}_0 := 1$.
FOR $n = 1, 2, \ldots$:
Forecaster announces $m_n \in [-1, 1]$.
Skeptic announces $M_n \in \mathbb{R}$.
Reality announces $x_n \in [-1, 1]$.
$\mathcal{K}_n := \mathcal{K}_{n-1} + M_n(x_n - m_n)$.

Intuitively, m_n is Forecaster's expected value of Reality's moves x_n. This means that he is prepared to sell to Skeptic, for m_n each, any real number of tickets (positive, zero, or negative) that pay x_n. The number of tickets Skeptic chooses to buy is M_n, and his capital at the end of round n of the game is $\mathcal{K}_n$ (assuming his initial capital was 1).

Let us say that a strategy for Skeptic is *prudent* if Skeptic's capital $\mathcal{K}_n$ never becomes negative, no matter what moves Forecaster and Reality make.

Theorem 1 *Skeptic has a prudent strategy in the Bounded Forecasting Game such that he becomes "exponentially rich", in the sense that*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathcal{K}_n > 0, \tag{3}$$

[1]The crucial step in developing that framework itself was adapting Schnorr's martingale definition of randomness deficiency to respect Dawid's [6, 7] prequential principle. As soon as the picture became clear, the randomness deficiency was removed, and the references to the algorithmic probability literature became out of place and were edited out.

on the paths where

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (x_i - m_i) = 0 \tag{4}$$

is violated.

Proof Without loss of generality we assume $m_n \equiv 0$. Skeptic can just use the following strategy $\mathcal{P}$ (described in [29], p. 69; this whole proof is a reversed version of the proof of Proposition 3.2 in [29]). For every real number ϵ, let $\mathcal{P}_\epsilon$ be the strategy that always buys $\epsilon\alpha$ tickets, where α is the current capital. The strategy $\mathcal{P}$ is to divide the initial capital of 1 into two sequences of accounts, A_k^+ and A_k^-, $k = 1, 2, \ldots$, with initial capital 2^{-k-1} in accounts A_k^+ and A_k^-, and then to apply $\mathcal{P}_\epsilon$ with $\epsilon := 2^{-k}$ to each account A_k^+, and to apply $\mathcal{P}_\epsilon$ with $\epsilon := -2^{-k}$ to each account A_k^-.

Suppose (4) is violated, e.g.,

$$\frac{1}{n} \sum_{i=1}^{n} x_i > \delta$$

for some $\delta > 0$ and for infinitely many n. Taking k so large that $\epsilon := 2^{-k}$ satisfies $\epsilon < \delta/2$, we can see that, for infinitely many n,

$$\frac{1}{n} \sum_{i=1}^{n} x_i > 2\epsilon.$$

In conjunction with $x_i^2 \le 1$, this implies

$$\epsilon \sum_{i=1}^{n} x_i - \epsilon^2 \sum_{i=1}^{n} x_i^2 > \epsilon^2 n;$$

using the inequality $t - t^2 \le \ln(1+t)$ (true for $t \ge -1/2$), we further obtain

$$\sum_{i=1}^{n} \ln(1 + \epsilon x_i) > \epsilon^2 n.$$

Since the capital $\mathcal{K}_n^{\mathcal{P}_\epsilon}$ attained by the strategy $\mathcal{P}_\epsilon$ is

$$2^{-k-1} \prod_{i=1}^{n} (1 + \epsilon x_i),$$

we can see that

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathcal{K}_n^{\mathcal{P}_\epsilon} > 0,$$

which, of course, implies (3).

The other case where (4) is violated, *viz.*

$$\frac{1}{n}\sum_{i=1}^{n} x_i < -\delta$$

for some $\delta > 0$ and for infinitely many n, is considered analogously: take k so large that $\epsilon := -2^{-k}$ satisfies $|\epsilon| < \delta/2$. ∎

Theorem 1 corresponds to Schnorr's [25] variant of the strong law of large numbers mentioned above. Analogous assertions can be proven for the law of the iterated logarithm and the recurrence property; the guaranteed speed of growth of Skeptic's capital when these laws are violated will be, of course, much lower.

Even if we are correct that algorithmic randomness is valuable mainly as a tool for discovery, this would justify its being taught to future researchers much more widely than it is now.

Acknowledgments

This article draws on our recent book [29] and on an article by Vovk [36]. We thank Laurie Snell and Vladimir V'yugin for useful discussions. Our research for the article was partially supported by EPSRC grant GR/R46670/01, BBSRC grant 111/BIO14428, EU grant IST-1999-10226, and NSF grant SES-9819116.

References

[1] Eugene A. Asarin. Some properties of Kolmogorov δ-random finite sequences. *Theory of Probability and its Applications*, 32:507–508, 1987.

[2] Eugene A. Asarin. On some properties of finite objects random in the algorithmic sense. *Soviet Mathematics Doklady*, 36:109–112, 1988.

[3] Eugene A. Asarin. *On some properties of finite objects random in the algorithmic sense*. PhD thesis, Moscow State University, 1988. Academic advisor: Andrei N. Kolmogorov.

[4] Jacob Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713.

[5] Alonzo Church. On the concept of a random sequence. *Bulletin of American Mathematical Society*, 46(2):130–135, 1940.

[6] A. Philip Dawid. Statistical theory: the prequential approach. *Journal of the Royal Statistical Society* A, 147:278–292, 1984.

[7] A. Philip Dawid. Calibration-based empirical probability (with discussion). *Annals of Statistics*, 13:1251–1285, 1985.

[8] J. L. Doob. *Stochastic Processes*. Wiley, New York, 1953.

[9] Peter Gács, J. Tromp, and Paul Vitányi. Algorithmic statistics. *IEEE Transactions on Information Theory*, 47(6):2443–2463, 2001.

[10] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin, 1933. English translation (1950): *Foundations of the theory of probability.* Chelsea, New York.

[11] Andrei N. Kolmogorov. Теория вероятностей и ее применения. In Математика и естествознание в СССР, pages 51–61. GONTI, Moscow and Leningrad, 1938.

[12] Andrei N. Kolmogorov. Вероятность. In Большая Советская Энциклопедия, volume 7, pages 508–510. Bol'shaya Sovetskaya Ehntsiklopediya, Moscow, 2nd edition, 1951.

[13] Andrei N. Kolmogorov. Теория вероятностей. In Математика, ее содержание, методы и значение, volume 2, pages 252–284. Izdatel'stvo AN SSSR, Moscow, 1956.

[14] Andrei N. Kolmogorov. Теория вероятностей. In Математика в СССР за сорок лет, volume 1, pages 781–795. Fizmatgiz, Moscow, 1959.

[15] Andrei N. Kolmogorov. On tables of random numbers. *Sankhya. Indian Journal of Statistics A*, 25(4):369–376, 1963.

[16] Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–7, 1965.

[17] Andrei N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions of Information Theory*, IT-14:662–664, 1968.

[18] Andrei N. Kolmogorov. Combinatorial foundations of information theory and the calculus of probabilities. *Russian Mathematical Surveys*, 38(4):29–40, 1983.

[19] Andrei N. Kolmogorov. On logical foundations of probability theory. In Yu. V. Prokhorov and K. Itô, editors, *Probability Theory and Mathematical Statistics*, volume 1021 of *Lecture Notes in Mathematics*, pages 1–5. Springer, 1983.

[20] Andrei N. Kolmogorov and Boris V. Gnedenko. Теория вероятностей. In Математика в СССР за тридцать лет, pages 701–727. Gostekhizdat, Moscow and Leningrad, 1948.

[21] Leonid A. Levin. On the notion of a random sequence. *Soviet Mathematics Doklady*, 14:1413, 1973.

[22] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, New York, 2nd edition, 1997.

[23] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[24] Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*. F. Deuticke, Leipzig and Vienna, 1931.

[25] C. P. Schnorr. Klassifikation der Zufallsgesetze nach Komplexität und Ordnung. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 16:1–21, 1970.

[26] C. P. Schnorr. A unified approach to the definition of random sequences. *Math Systems Theory*, 5:246–258, 1971.

[27] C. P. Schnorr. *Zufälligkeit und Wahrscheinlichkeit*. Springer, Berlin, 1971.

[28] C. P. Schnorr. Process complexity and effective random tests. *Journal of Computer and System Sciences*, 7(4):376–388, 1973.

[29] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, New York, 2001.

[30] Alexander Shen'. The concept of Kolmogorov (α, β)-stochasticity and its properties. *Soviet Mathematics Doklady*, 28:295–299, 1983.

[31] Ray J. Solomonoff. A preliminary report on a general theory of inductive inference. Technical Report ZTB-138, Zator Company, Cambridge, MA, November 1960.

[32] Ray J. Solomonoff. A formal theory of inductive inference. Parts I and II. *Information and Control*, 7:1–22 and 224–254, 1964.

[33] Nikolai Vereshchagin and Paul Vitányi. Kolmogorov's structure functions with an application to the foundations of model selection. In *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 2002. To appear.

[34] Jean Ville. *Etude critique de la notion de collectif*. Gauthier-Villars, Paris, 1939.

[35] Vladimir Vovk. The law of the iterated logarithm for random Kolmogorov, or chaotic, sequences. *Theory of Probability and its Applications*, 32:413–425, 1987.

[36] Vladimir Vovk. Kolmogorov's complexity conception of probability. In Vincent F. Hendricks, Stig Andur Pedersen, and Klaus Frovin Jø rgensen, editors, *Probability Theory: Philosophy, Recent History and Relations to Science*, pages 51–69. Kluwer, Dordrecht, 2001.

[37] Vladimir V. V'yugin. On nonstochastic objects. *Problems of Information Transmission*, 21:3–9, 1985.

[38] Vladimir V. V'yugin. On the defect of randomness of a finite object with respect to measures with given compexity bounds. *Theory of Probability and its Applications*, 32:508–512, 1987.

[39] Abraham Wald. Die Widerspruchfreiheit des Kollectivbegriffes der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Mathematischen Kolloquiums*, 8:38–72, 1937.