Труды Москов. Матем. Общ.
Том 70 (2009)

Trans. Moscow Math. Soc.
2009, Pages 31–69
S 0077-1554(09)00180-0
Article electronically published on December 3, 2009

STOCHASTIC AND DETERMINISTIC CHARACTERISTICS OF ORBITS IN CHAOTICALLY LOOKING DYNAMICAL SYSTEMS

V. I. ARNOLD

Abstract. We study finite length sequences of numbers which, at the first glance, look like realizations of a random variable (for example, sequences of fractional parts of arithmetic and geometric progressions, last digits of sequences of prime numbers, and incomplete periodic continuous fractions).

The degree of randomness of a finite length sequence is measured by the parameter introduced by Kolmogorov in his 1933 Italian article published in an actuarial journal.

Unexpectedly, fractional parts of terms of a geometric progression behave much more randomly than terms of an arithmetic progression, and the statistics of periods of continuous fractions for eigenvalues of unimodular matrices turns out to be different from the classical Gauss–Kuzmin statistics of partial continuous fractions of random real numbers.

Empirically, the lengths of the period of continuous fractions for the roots of quadratic equations with leading coefficient 1 and increasing other (integer) coefficients, grow, on the average, as the square root of the discriminant of the equation.

1. Kolmogorov's stochasticity parameter

Laplace was calculating the probability that the sun will rise tomorrow. It is not easy to determine what he meant by "probability".

First attempts to define an objectively measurable degree of randomness of observable events were made by von Mises and Kolmogorov. In his 1933 Italian paper on actuarial mathematics (see Kolmogoroff [1]), Kolmogorov introduced his "stochasticity parameter" λ_n of a set of n real numbers, which allows one to estimate how realistic is the assumption that these n numbers are n independent values of the same real random variable.

The definition of this stochasticity parameter λ_n begins with the description of the given set of n real numbers using the "empirical counting function" C_n defined below.

Let us arrange the given numbers in increasing order,

$$x_1 \le x_2 \le \cdots \le x_n.$$

The value of the *empirical counting function* $C_n(X)$ is defined as the number of elements x_m that are not larger than X:

$$\begin{aligned} C_n(X) &= 0 \quad \text{for } X < x_1; \\ C_n(X) &= m \quad \text{for } x_m \le X < x_{m+1}; \\ C_n(X) &= n \quad \text{for } x_n \le X. \end{aligned}$$

Kolmogorov assumed that the random variable $x \in \mathbb{R}$ has a continuous distribution function and describes the "*theoretical counting function*" $C_0(X)$ as the expectation of

2000 *Mathematics Subject Classification.* Primary 37A45.

The work was partially supported by RFFI, Grant 05-01-00104.

©2009 American Mathematical Society

the number of values not exceeding X among n independent observations of the random variable x:

$$C_0(X) = n \cdot (\text{probability that } x \le X).$$

The empirical counting function C_n differs from the theoretical one, and Kolmogorov measures their divergence using the uniform norm

$$\mathcal{F}_n = \sup_X |C_n(X) - C_0(X)|.$$

Viewing both C_n and C_0 as sums of approximately n more or less independent random variables (with zero expectation for the difference $C_n - C_0$), one can easily see that the values of the random variable $\mathcal{F}_n$ are of order $\sqrt{n}$ (following the theory of Brownian motion).

Therefore, Kolmogorov normalizes the difference and defines his *stochasticity parameter* (of the given sequence $x_1 \le x_2 \le \cdots \le x_n$ of real numbers) by the formula

$$\lambda_n = \frac{\mathcal{F}_n}{\sqrt{n}}\,.$$

To avoid using the theoretical distribution (i.e., the function C_0), one can replace this function C_0 in the definition of the divergence $\mathcal{F}_n$ with the empirical counting function C'_n constructed using another sample of n values. Presumably, the resulting modified stochasticity parameter is asymptotically proportional to λ_n, but I do not know the exact value of the proportionality coefficient (or its distribution). Empirically it seems to be about 30% higher than Kolmogorov's λ_n, but can take values between λ_n and $2\lambda_n$.

One can probably find more about this from the article [27] by N. V. Smirnov.

The Kolmogorov stochasticity parameter λ_n is itself a random variable with the *distribution function* Φ_n:

$$\Phi_n(\Lambda) = (\text{probability that } \lambda_n \le \Lambda).$$

Kolmogorov proved the following result.

Theorem. *As $n \to \infty$, the distribution functions Φ_n tend (uniformly) to the universal limit Φ (independent of the continuous distribution function C_0/n of the random variable x whose n values ($x_1 \le \cdots \le x_n$) were used to compute the values of the parameter λ_n and their distribution function Φ_n).*

Remark. The fundamental universality property (independence of Φ of C_0) has a simple relativistic origin: any two continuous probability distributions are equivalent in the sense that they differ only by a choice of the coordinate on the x axis. Therefore every characterization of the distribution defined independently of the coordinate x will take the same value on any two continuous probability distributions. In particular, the distributions Φ_n (hence also Φ) of the values of the stochasticity parameter Λ do not depend on the choice of the coordinate on the x axis, hence are universal.

This remark allows one to compute these distributions explicitly: it suffices to consider the uniform distribution C_0 for which the computation of the probability $\Phi_n(\Lambda)$ of the even $\lambda_n \le \Lambda$ reduces to the summation of volumes of certain simplices in the Euclidean space.

The resulting formulas for the distributions Φ_n contain many factorials, but the Stirling asymptotic formula yields the short Kolmogorov formula for the universal distribution Φ. Namely, Kolmogorov proved that for $\Lambda > 0$ we have

$$\Phi(\Lambda) = \sum_{k=-\infty}^{k=+\infty} \Big((-1)^k e^{-2k^2\Lambda^2} \Big).$$

This *Kolmogorov distribution function* smoothly increases from 0 (at $\Lambda = 0$) to 1 (as $\Lambda \to +\infty$), remaining very small for, say, $\Lambda \le 0.4$ (for $\Lambda \to 0$ its asymptotics is

$$\Phi(\Lambda) \sim \frac{\sqrt{2\pi}}{\Lambda}\, e^{-\frac{\pi^2}{(8\Lambda^2)}},$$

so that all derivatives of Φ at the point $\Lambda = 0$ equal 0).

The values of the function $\Phi(\Lambda)$ are close to 1 for, say, $\Lambda \ge 1.8$. In [1], Kolmogorov presented the following table of Φ with four-digit precision:

Λ	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
$10^4\Phi$	28	1357	4558	7300	8877	9603	9888	9969

The average of the stochasticity parameter Λ distributed according to this law equals $\Lambda = \sqrt{\pi/2}\ln 2 \approx 0.867$, and the median equals $\Lambda_* \approx 0.83$ (the events $\Lambda \le \Lambda_*$ and $\Lambda \ge \Lambda_*$ both have probability 50 %).

The idea of the application of this Kolmogorov theorem to measuring the degree of randomness is as follows. First we compute the empirical value λ_n (for some sample of n observed values of the variable x whose randomness we study).

If the resulting value λ_n is significantly smaller than the average $\Lambda \approx 0.87$, then the probability $\Phi(\lambda_n)$ of the event $\Lambda \le \lambda_n$ is so small that x is unlikely to be random.

For example, for the empirical value $\lambda_n = 0.4$ the probability $\Phi(0.4) \approx 28/1000$ of the appearence of such a small value of the stochasticity parameter for n independent observations of a random variable x is less than one third of one per cent, and the experimenter should conclude that x is unlikely to be random.

Equally unlikely is the appearance of λ_n that is much larger than the average Λ. For example, if $\lambda_n = 1.8$, then $\Phi(\lambda_n) \approx 0.9969$. Therefore the probability of the appearance of this (or larger) value of the stochasticity parameter (for a sample of n values of a random variable x) is less than one third of one per cent, and the experimenter should again conclude that x is unlikely to be random.

The estimate of the objective degree of randomness by von Mises did follow the same scheme but with a different choice of metric used to estimate the divergence $\mathcal{F}_n$ of counting functions: instead of the uniform norm, von Mises used the quadratic mean, i.e., the Hilbert norm.

One of the best applications of the Kolmogorov stochasticity parameter was described by Kolmogorov himself in 1940. Unfortunately, not all of Kolmogorov's conclusions were published (for political reasons).

At that time, followers of T. D. Lysenko, who was fighting against classical genetics, published the results of their experiment refuting Mendel's Law of Independent Assortment (explaining the 3 : 1 ratio for the appearance of white and purple flowers on pea plants in the second generation).

In the described experiment, out of 4000 plants the observed ratio was something like 940 : 3060 instead of the theoretical ratio 1 : 3. Lysenko concluded that this difference refutes the Mendel law.

Computing his stochasticity parameter for these experimental data, Kolmogorov obtained the value $\Lambda \approx 0.8$, confirming the honesty of the experimenters.

Kolmogorov also came to the conclusion that if, in another experiment, the result would have been closer to Mendel's 3 : 1 prediction, then in the experiment of Lysenko's followers, this would have indicated that the data were falsified: any difference smaller than $1000 - 940 = 60$ is unlikely.

However, it was impossible to publish the conclusion in this form because, at the very same moment, classical geneticists submitted an article (to the same journal, Doklady

AN SSSR) that refuted the work of Lysenko's pupils: experimental results reported in this article were very close to Mendel's prediction.

Kolmogorov's conclusion would mean that these results are falsified, as opposed to the results of Lysenko's pupils who presented an honest report about their experiment. Not willing to write about a falsification, Kolmogorov published, in the article about a new confirmation of Mendel's laws (see [2]), just his report on Lysenko's attempt to refute Mendel's laws. Even later, Kolmogorov thought that Lysenko was just an illiterate boor.

The Kolmogorov stochasticity parameter can be used to estimate the degree of randomness of many other pseudorandom sequences. Being more a method of natural sciences than a mathematical theorem, the comparison of empirically found values of the Kolmogorov stochasticity parameter λ_n with its distribution function Φ can be useful both to establish nonrandomness when the empirical values of λ_n are far from 0.87 and to confirm randomness when $|\lambda_n - 0.87|$ is small.

Whichever the source of observed "chaotic behavior" is, the objective conclusion that the stochasticity is too small (for example, the observation that $\lambda_n \to 0$ as $n \to \infty$) can provide useful information about the nature of the variable we are considering.

The observation that the values of the stochasticity parameter λ_n for samples we are studying stay in the region where $\Phi(\lambda_n)$ is far away both from 0 and from 1, can be viewed as an "objective confirmation of randomness" (regardless of the origin of sequences of numbers we are studying; as we will see later, these sequences can be provided by completely deterministic dynamical systems or by simple number-theoretic algorithms).

For example, both sequences below, each formed by 15 two-digit integers look similarly random; however, the values of the Kolmogorov stochasticity parameter for these two sequences are sharply different. The sequence of remainders modulo 100 of terms of a geometric progression

$$03,\ 09,\ 27,\ 81,\ 43,\ 29,\ 87,\ 61,\ 83,\ 49,\ 47,\ 41,\ 23,\ 69,\ 07 \tag{1.1}$$

turns out to be "objectively chaotic" (in this sense), whereas the sequence of remainders modulo 100 of terms of an arithmetic progression

$$37,\ 74,\ 11,\ 48,\ 85,\ 22,\ 59,\ 96,\ 33,\ 70,\ 07,\ 44,\ 81,\ 18,\ 55 \tag{1.2}$$

turns out to be "objectively nonrandom".

Towards the end of his life, Kolmogorov developed a new axiomatic approach to probability that uses von Mises's suggestion to consider as random any sequence of numbers "whose continuation cannot be guessed".

Von Mises did not define precisely the words "cannot be guessed", whereas Kolmogorov used the mathematical logic approach requesting that "there exists no finite algorithm, i.e., a finite program for a Turing machine, predicting the continuation of our sequence".

Kolmogorov proved that from this new definition of randomness one can deduce many classical theorems of probability, mathematical statistics, and information theory. However, in all problems discussed earlier in this paper, this approach is useless since it only works when the number n of trials grows indefinitely; from this asymptotic point of view, *all finite sequences of numbers have zero randomness* since any finite sequence can be generated by a Turing machine with a finite program.

On the contrary, the stochasticity parameter λ_n allows us to compare "the degree of randomness" of finite sequences (1.1) and (1.2):

$$\lambda_{15}(1) \approx 0.70, \qquad \Phi(0.70) \approx 0.30;$$
$$\lambda_{15}(2) \approx 0.33, \qquad \Phi(0.33) < 0.001.$$

In that sense (in the sense of the degree of randomness) the first sequence (1.1) is at least 300 times more random than the second sequence (1.2).

On the other hand, from the algorithmic point of view, both sequences have equally simple origins: (1.1) is the sequence of modulo 100 residues of the geometric series (with ratio $a = 3$)

$$(a^x \pmod{100}) \qquad (x = 1, 2, \ldots, (n = 15))$$

and (1.2) is the sequence of modulo 100 residues of the arithmetic series (with difference $b = 37$)

$$(bx \pmod{100}) \qquad (x = 1, 2, \ldots, (n = 15)).$$

As we will see, geometric progressions are much more random than arithmetic progressions.

2. Applications of the degree of randomness to particular sequences

Kolmogorov proved his theorem for real random variables (with continuous distribution functions). Earlier I (unlawfully) applied Kolmogorov's universal distribution Φ to variables assuming only integer values or even just a finite numbers of values (remainders of division by an integer N).

I know, of course, that mathematical rigor does not allow us to make unsubstantiated generalizations like that. But as a natural scientist I believe that, e.g., results in astronomy should not depend on whether the distance measured in some units (be it kilometers, miles, angstroms, or parsecs) takes real or just integer values.

Therefore I hope that one can apply the Kolmogorov theory not just to real random variables (with continuous distribution functions) but also to other variables; for example, this theory should have generalizations to random variables whose values are integers $x \in \mathbb{Z}$, or points on the circle S^1, or remainders $x \in \mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$, or even rational numbers ($x \in \mathbb{Q}$).

Of course, all these generalized theories should be rigorously formulated and proofs should be given, I hope this will be done (by mathematicians of the future). However, even before that I will be using Kolmogorov's distribution Φ in these more general cases (in the hope that it is a sufficiently good approximation to genuine distributions of the randomness parameter in these generalized theories).

Example 2.1. Modulo N remainders of n terms

$$\{a^x A \pmod{N}\} \qquad (n = 0, 1, \ldots, (n-1)) \tag{2.1}$$

of the geometric progression with the integer first term A and integer ratio $1 < a < N$ can look like a random sample of points uniformly distributed in $\mathbb{Z}_N$ provided that the number n of terms is not "too large". For example, we can take $n \approx T/2$ or θT with the constant θ separated from 0 and from 1, $0 < \theta < 1$, where $T = T(N, a)$ is the period of the sequence (2.1) consisting of remainders of terms of progression modulo N.

For different initial points A (of the dynamical system $\mathbb{Z}_N \to \mathbb{Z}_N$ sending x to ax) the Kolmogorov stochasticity parameter λ_n of the corresponding n remainders of terms of the geometric progression takes different values $\lambda_n(A)$.

Computing all these numbers and counting their distribution (corresponding to the uniform distribution of the initial point A on $\mathbb{Z}_N$) I have found (based on several hundreds of such experiments) a reasonable similarity of the distribution of obtained values $\lambda_n(A)$ of the Kolmogorov stochasticity parameter for different orbits of our dynamical system (i.e., for different values of A) with the universal distribution Φ of the stochasticity parameter.

In the majority of examples the length of the progression was taken to be $n = 50$ (to simplify the computation of remainders of division by $N = 100$). The similarity with the Kolmogorov distribution Φ is not a theorem but an empirical observation. In

mathematical terms, it should be called a "conjecture" that as $N \to \infty$, the distribution of the values of the Kolmogorov parameter for mod N remainders of terms of N geometric progression (corresponding to N initial points A) tends to Φ.

Of course, my experiments (which yield at most several million values of $\lambda_n(A)$) do not prove this conjecture. Nevertheless, I believe it is true.

The "theoretical distribution" C_0 of mod N remainders in these experiments was assumed to be uniform (i.e., each of the N remainders modulo the prime number N was assumed to have the same probability $1/N$). As far as I know, this conjecture about the uniform distribution of remainders is not yet rigorously proved,[1] but in the book [3] (about Galois theory) I gave some "physical proofs" of (more general) theorems about uniform distribution (including the uniform distribution of fractional parts of numbers a^x for almost all real bases a (and the set of exceptional values of a has Lebesgue measure 0).

These "physical arguments" are not proofs in the mathematical sense because they use "adiabatic approximations" which are often used in physics but do not have, so far, rigorous mathematical proofs. However, I hope that these statements about the uniform distribution are correct and, moreover, that the necessary justifications of adiabatic theorems will eventually be rigorously proved.

The empirical study of various (exponentially growing) recurrent sequences, including the Fibonacci sequence, leads to similar conclusions about the uniform distribution of remainders that are similar to those in the above case of geometric progression.

Example 2.2. A long geometric progression with the integer ratio a, where $1 < a < N$,

$$\{a^x A \pmod N\} \qquad (x = 0, 1, \dots, n-1),$$

consisting of remainders modulo a fixed integer N (for example in the case where N is a prime number).

Theorem. *The Kolmogorov stochastisity parameter λ_n of such a sequence of n remainders of division by N tends to* 0 *as the number n of remainders tends to infinity.*

This follows from the (small) Fermat theorem about the periodicity of the sequence of remainders in the case where N is a prime, or from Euler's generalization of this theorem: Euler proved that the sequence of remainders of division by any integer is periodic.

To determine the stochasticity parameter in the latter case one should either assume that the base a is relatively prime to N or correct the "uniform distribution" assuming it to be uniform only on those remainders of division by N that are divisible by all common divisors of a and N.

Now we consider a more general case where the base a is not assumed to be an integer but is allowed to be an arbitrary real number $a > 1$.

Conjecture. *The Kolmogorov stochasticity parameter λ_n of residues modulo N of n terms of a geometric progression with an arbitrary ratio $a > 1$ does not tend to zero as $n \to \infty$ (for almost all a, so that exceptional values of a form a set of Lebesgue measure zero on the real line).*

Moreover, one can conjecture that for almost any base $a > 1$ the following more general statement holds (see [3]):

The distribution of the values $\lambda_n(A)$ of the Kolmogorov stochasticity parameter λ_n of the sequences of n remainders modulo N of geometric progressions starting at different points A $(0 < A < N)$, tend, as $n \to \infty$, to the universal Kolmogorov distribution Φ

[1] I am grateful to A. A. Karatsuba who brought to my attention the article by J. F. Koksma, *Ein mengentheoretischer Satz über die Gleichverteilung modulo Eins*, Compositio Math. **2** (1935), 250–258.

(under the assumption that the starting point A is uniformly distributed on the interval $0 < A < N$).

Now let us consider remainders modulo N of terms of *arithmetic progressions*. To simplify the forthcoming exposition we choose the scale taking N to be equal to 1, so that our sequences consist of *fractional parts*

$$\{ax\} \qquad (x = 1, 2, \dots, n).$$

Every real number z is represented as the sum $z = [z] + \{z\}$ of the integral part $[z] \in \mathbb{Z}$ and the fractional part $\{z\}$, $0 \le \{z\} < 1$.

Theorem (see [5]). *The Kolmogorov stochasticity parameter λ_n of n fractional parts of terms of an arithmetic progression with a rational difference a tends to a as $n \to \infty$.*

This implies that sequences formed by remainders of division by N of terms of long arithmetic progressions with the difference a commensurable with N (this always takes place if, for example, both N and a are integers) is never random (provided that the degree of randomness is measured by the Kolmogorov parameter λ_n).

However, this asymptotic vanishing of λ_n (as $n \to \infty$) does not occur, for example, for some irrational differences of the corresponding arithmetic progressions: it is proved in [5] that *there exist (irrational) numbers a such that the sequence of values λ_n of the Kolmogorov stochasticity parameter for fractional parts of n terms of the arithmetic progression with difference a does not tend to* 0 *as $n \to \infty$. Moreover, there exist values of a for which this sequence contains an infinite number of arbitrarily large elements $\lambda_{n_j} \ge K$.*

These irrational differences a form, similarly to rational values of a, an everywhere dense set. The appearance of arbitrarily large values $\lambda_n \ge K$ follows from some properties of incomplete continuous fractions of the number a. Therefore *every asymptotically defined property* of the sequence of values of the stocasticity parameter for remainders of n terms of an arithmetic progression (for example "$\lambda_n \to 0$", or "$\lambda_n \not\to 0$", or "$\lambda_n \to \infty$", or "$\lambda_n \ge K$ infinitely many times", etc.) *either occurs almost always* (so that the exceptional values of a have measure 0) or *occurs almost never* (i.e., the set of those a for which this property holds has measure 0).

Unfortunately, I don't know which alternative ("almost always" or "almost never") holds for the properties formulated above: whether remainders of almost all arithmetic progressions are random or nonrandom as far as the behavior of the values λ_n of the stocasticity parameter of the first n elements of the sequence is concerned.

This general question is difficult to check both theoretically and experimentally: an empirical study of the fractional parts of arithmetic progressions presumably requires answering nontrivial questions about the statistics of continuous fractions, and the standard "Gauss–Kuzmin" statistics describing the distribution of incomplete continuous fractions of random real numbers (and their finite combinations) is insufficient to solve the above nontrivial problems.

There is another version of these asymptotic problems, which be easier to solve. It turns out that unusually large values λ_n of the stochasticity parameter of the sequence of n remainders of terms of an arithmetic progression occur (for many values of the difference a) only for some rare special values of the length n. To avoid the dominating influence of these rare events one can replace the values λ_n with their Cesàro means:

$$\widehat{\lambda}_n = \frac{\lambda_1 + \dots + \lambda_n}{n}.$$

Quite irregular dependence of the values of $\lambda_n(a)$ on arithmetical properties of the difference a of the corresponding arithmetic progression can be smoothened by passing

from $\lambda_n(a)$ to the average values $\widehat{\lambda}_n(a)$, whose dependence on arithmetical properties of a can be easier to investigate.

Another averaging method is provided by the theory of *weak asymptotics*; see [6].

For a rational difference $a = p/q$ we replace the sequence of fractional parts of the arithmetic progression by the sequence of remainders of division by q of terms of the progression with difference p (this sequence has the same stochasticity parameter $\lambda_n(p,q) = \lambda_n(a)$ for each n).

For each value M of the scaling factor we consider the radius r of the neighborhood U of the point $(Mp, Mq) \in \mathbb{R}^2$. For each integral point (P', Q') in this neighborhood we consider the value of the stochasticity parameter $\lambda_n(a)(P', Q')$ and compute the average $\widehat{\lambda}_n^*$ of these real numbers over all integral points in the neighborhood U.

As the scaling factor M tends to infinity, the values $\widehat{\lambda}_n^*$ can have a limit $\widehat{\lambda}_n^*(a)$, which is then called the *weak asymptotic of* $\lambda_n(a)$.

It is important that, in taking the limit, the number n of the terms of the progression remains fixed; indeed, the values of the stochasticity parameter $\lambda_n(P', Q')$ tend to 0 as $n \to \infty$ whereas their averages $\widehat{\lambda}_n^*$ may behave differently!

Here the "theoretical distribution" C_0 used to determine the values of the stochasticity parameters λ_n was chosen to be uniform: each of the N remainders of division by N was assigned (in the integral case we consider here) probability $1/N$ provided the difference a of the progression is relatively prime with N.

For a progression with a rational difference, the fact that long sequences of remainders are uniformly distributed is easy to prove. However, the uniform distribution of remainders is true for progressions with irrational differences as well (as was proved by H. Weyl).

Example 2.3. As yet another example of a quasirandom sequence, we consider remainders of division by some integer N of the *sequence of prime numbers*. Take, for example the following sequence with $n = 21$ terms:

$$01, 03, 07, 13, 27, 31, 37, 39, 49, 51, 57, 63, 67, 73, 79, 81, 91, 93, 97, 99,$$

consisting of modulo $N = 100$ remainders of prime numbers $(101, 103, \ldots, 199)$ between 100 and 200.

In [5], I compared the counting function C_{21} of this sequence with the "theoretical distribution" C_0 based on the "density of the distribution of prime numbers" found by Legendre and Chebyshev (which, near n, is inversely proportional to $\ln n$).

The resulting value of the Kolmogorov stochasticity parameter turns out to be approximately

$$\lambda_{21} \approx 0.5, \qquad \Phi(0.5) \approx 0.07.$$

This shows that the distribution of prime numbers between 100 and 200 does not resemble a random distribution, so that the probability of its being random computed using the Kolmogorov stochasticity parameter is approximately 7%.

This observation, which contradicts the opinion of many specialists in number theory, can be explained as follows. Near an integer n, the average distance between two consecutive prime numbers is $\ln n$ (according to the Legendre distribution). For large n, the function $\ln n$ grows slowly because the derivative $\frac{d \ln n}{dn} = \frac{1}{n}$ is small. Therefore, the sequence of prime numbers looks like an arithmetic progression (with "slowly growing" difference).

This similarity allows one to explain the "observed nonrandomness" of remainders of prime numbers by similar nonrandomness of remainders of arithmetic progressions. A weak point of this explanation is that the smallness of the stochasticity parameter λ_n

for remainders of division by N of terms of an arithmetic progression is proved only in the case where the difference of the progression is commensurable with N. On the other hand, the "slowly growing" difference $\ln n$ in the Legendre distribution takes values a both commensurable and noncommensurable with $N = 100$.

Empirically, it seems that longer (e.g., consisting of $n = 500$ terms) sequences of consecutive prime numbers starting, for example, with the first prime number exceeding one million, are more random. However, there are no known theorems about the behavior, for large n, of values of the stochasticity parameter λ_n for remainders of division by N of n consecutive primes.

3. Statistics of continuous fractions

A *continuous fraction with (natural) elements* $(a_0, a_1, \dots, a_n)$ is a rational number given by the formula

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n}}};$$

this number is also denoted $[a_0, a_1, \dots, a_n]$. The number $a_0 = [x]$ is allowed to be negative, but $a_j > 0$ for $j > 0$.

Similarly one can consider infinite continuous fractions $[a_0, a_1, \dots]$; the value of such a fraction is defined as the limit, as $n \to \infty$, of truncated fractions

$$\frac{p_n}{q_n} = [a_0, a_1, \dots, a_n].$$

Example. The "golden ratio"

$$x = \frac{\sqrt{5}+1}{2}$$

can be expanded in the infinite continuous fraction $[1, 1, 1, \dots]$.

Indeed, for this continuous fraction $x = 1 + \frac{1}{1+\cdots}$ we have $x = 1 + \frac{1}{x}$, so that

$$x^2 - x - 1 = 0, \qquad x = \frac{1}{2} + \sqrt{\frac{1}{4} + 1} = \frac{\sqrt{5}+1}{2} \approx 1.6\dots.$$

Truncated continuous fractions p_m/q_m of x yield very good rational approximations of the irrational number x since

$$\left| x - \frac{p_m}{q_m} \right| \le \frac{1}{q_m q_{m+1}}.$$

This approximation is particularly precise when q_{m+1} is large (it is always larger than q_m since $q_{m+1} = q_{m-1} + a_{m+1} q_m$).

Example. The fraction $\frac{p}{q} = \frac{355}{113}$ (obtained by slicing the number 113 355) yields a very good approximation of the number $\pi \approx 3.14159265\dots$ with 6 correct digits ($p/q \approx 3.1415929\dots$).

The fraction p/q is obtained using the continuous fraction

$$\pi = [3, 7, 15, 1, 292, \dots]$$

by cutting in front of the large element:

$$\frac{p}{q} = 3 + \cfrac{1}{7 + \cfrac{1}{16}}.$$

The approximation

$$\pi \approx 3 + \frac{1}{7} = \frac{22}{7}$$

used by Archimedes is good for the same reason (the next element 16 is large).

If the ratio of periods of evolution of two planets around the Sun is close to a rational number (in celestial mechanics this is called a *resonance*), then these two planets perturb each other stronger than in the nonresonance case.

For Jupiter and Saturn (passing approximatively 299 and 120.5 angular seconds a day along the orbit) the ratio of angular velocities $120.5/299 \approx 2/5$ is close to a resonance value. This resonance generates a "large secular perturbation" that increases for several hundreds of years (before the planets return to their previous orbits after about 900 years).

Thinking about the stability of the solar system, H. Poincaré suggested to investigate how many frequencies in it satisfy dangerous resonance relations, using for this purpose continuous fractions of ratios of frequencies: large elements of these continuous fractions generate strong resonances because then the corresponding ratio has a good rational approximation by the appropriately truncated continuous fraction. Therefore, it is important to know how many of those ratios have continuous fractions with large elements (having relatively small indices).

This brought up an important mathematical theory called the statistics of continuous fractions: this theory studies whether for a given real number all elements of its continuous fraction are small (as for the golden ratio, where they all are equal to 1), or some of them are large (which turned out to be typical for the majority of real numbers x).

This natural problem was also studied by "pure mathematicians", regardless of its applications in the natural sciences.

Gauss obtained the following expression for the "frequency" of appearance of a number k among elements of an "arbitrary" continuous fraction:

$$f_k = \frac{1}{\ln 2} \ln \frac{(k+1)^2}{k(k+2)} = \frac{1}{\ln 2} \ln\left(1 + \frac{1}{k(k+2)}\right). \tag{3.1}$$

Example. For small k, the frequency $f(k)$ is rather large (decreasing as $1/k^2$ as $n \to \infty$):

k	1	2	3	4	5	6	7	8
$100 f_k$	47	17	9	6	4	3	2	2

The difficulty of this theory is that Gauss not only didn't prove his statements, but didn't even formulate them explicitly. The justification of Gauss' results was obtained only by Kuzmin who proved in 1928 (in [7]) the following theorem.

Theorem. *For almost every real number x the frequency of k among the elements of the continuous fraction of x equals the number $f(k)$ given by the Gauss formula* (3.1).

The frequency in the Kuzmin theorem is defined as the following limit. Denote by $\mathcal{F}_n(k)$ the number of times k occurs among the first n elements of the continuous fraction expansion $x = [a_0, a_1, a_2, \dots]$ of the number x.

Kuzmin proved that

$$\lim_{n\to\infty} \frac{\mathcal{F}_n(k)}{n} = f_k \tag{3.2}$$

for almost all real numbers x (i.e., for all real numbers except some special ones forming a set of Lebesgue measure zero on the real axis x).

One of the most surprising properties of the statistics (3.2) is that the limit (3.2) does not depend on the real number x whose continuous fraction is described by this statistics: statistics (3.2) is *universal*.

Exclusion of some special numbers is necessary, and these special numbers form an everywhere dense set in the real line. Indeed, relation (3.2) fails, for example, for all rational numbers and also for many irrational numbers (including, in particular, the golden ratio).

Of course, Gauss knew about these exceptional cases where his statistics (3.1) fails, and these counterexamples did not allow him to publish his discovery. He could not even give a correct statement because the Lebesgue measure was not yet discovered, and without using the Lebesgue measure it was difficult to formulate a statement that the found statistics is generic (in Kuzmin's formulation this is expressed by the words "for almost all x").

Gauss made his discovery while studying an interesting dynamical system whose phase space is the interval (0,1) of the real line x:

$$g\colon (0,1) \to (0,1), \qquad g(x) = \{1/x\}$$

(as usual, by $[t]$ we denote the integral part of a real number x and by $\{t\} = t - [t]$, $0 \le \{t\} < 1$, the fractional part of x).

To avoid the appearance of the value $\{1/x\} = 0$ which makes an orbit of the dynamical system nonextendable, it suffices to exclude from the phase space all rational numbers (which form a set of Lebesgue measure zero in the phase space).

Gauss noticed that this dynamical system has an invariant measure

$$\mu(A) = \frac{1}{\ln 2} \int_A \frac{dx}{1+x}. \tag{3.3}$$

The factor $1/(\ln 2)$ in (3.3) is introduced to make the measure of the entire phase space $(0,1)$ to be equal to 1.

The invariance of the measure does not mean that $\mu(gA) = \mu(A)$: in the Gauss dynamical system this condition is not satisfied.

Similarly to differential forms, measures are carried by maps *backward* rather than forward. The *invariance* of the measure μ under the map g is the condition

$$\mu(g^{-1}A) = \mu(A) \tag{3.4}$$

(for each measurable set A).

Relation (3.4) holds for the Gauss measure (3.3) and to prove it is a good exercise. This proof led Gauss to the discovery of the method of "telescopic summation". For simplicity, I will explain this method here on the simplest example. Let us compute the sum of the series

$$S = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \frac{1}{3\cdot 4} + \cdots .$$

The method consists in using the identity

$$\frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)},$$

which shows that

$$S = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots = 1.$$

Writing explicitly the infinite sequence of intervals forming the complete preimage $g^{-1}(A)$ of the interval A and integrating the density (3.3) over this infinite sequence of intervals, we obtain for $\mu(g^{-1}(A))$ the series converging (telescopically) to $\mu(A)$.

Modern books about continuous fractions derive Kuzmin's theorem from the Birkhoff theorem about ergodic dynamical systems by applying it to the Gauss map g (see, e.g., the book [8]).

However, Kuzmin's proof was different. He found it before the Birkhoff theorem appeared (and knew nothing about it). I think that the original Kuzmin's proof contained at least a version of the Birkhoff theorem about the dynamical system Kuzmin needed (and it would be interesting to see whether his arguments are applicable in some more general case he didn't consider).

Nowadays we can use the Birkhoff theorem, forgetting about the preceding proof given by Kuzmin. The point is that our dynamical system increases the distance between points of the phase space because

$$\left|\frac{d(1/x)}{dx}\right| = \left|\frac{1}{x^2}\right| > 1 \quad \text{for } 0 < x < 1.$$

This increase of distance with time leads to the ergodicity (and chaoticity) of the Gauss dynamical system: a small perturbation of the initial point x results, over time, in a large perturbation of its image $g^t x$ (after sufficiently large motion time t).

According to ergodic theory, time averages of (measurable) functions coincide (for almost all initial points of orbits) with the average of this function over the entire phase space.

The event "$a_j = k$" for the j-th elements of the continuous fraction for x means that

$$\frac{1}{k+1} < y \le \frac{1}{k} \quad \left(\left[\frac{1}{y}\right] = k\right)$$

for the j-th image y of the initial point x.

Computing the measure μ of the interval given by formula (3.3) we obtain the Gauss expression (3.1):

$$\mu\left(\left(\frac{1}{k+1}, \frac{1}{k}\right)\right) = \frac{1}{\ln 2}\left(\ln\left(\frac{1}{1+\frac{1}{k}}\right) - \ln\left(\frac{1}{1+\frac{1}{(k+1)}}\right)\right) = f_k$$

for the probability of the event $a_j = k$ (hence, by the ergodic theorem, for the required frequency of the number k among the elements of the continuous fraction of a random number x).

Long before Kuzmin's paper appeared, the practical consequences of his theory were studied empirically.

Poincaré suggested to put together lists of continuous fractions for the ratios of various periodic motions observed in the solar system (like ratios of the year lengths of various planets and month lengths of various satellites) in order to find the statistics of resonance cases. He thought that the existence of too many too strong resonances can make the planet system unstable, whereas the rare occurence of strong resonances could indicate probable stability of the motion of planets for millions of years of the existence of the planetary system.

Resulting tables of continuous fractions observed in astronomy were published in 1888 by Poincaré's follower Gylden; see [9].

The observed statistics of these continuous fractions turned out to be very similar to the Gauss–Kuzmin statistics [1].

These experimental facts do not have a scientific explanation: the corresponding continuous fractions depend on the initial positions of planets in the system, and the fact that today's "initial" conditions satisfy the Gauss–Kuzmin statistics is not explained (by any presently known theory of the formation of planetary systems).

Despite the lack of analysis of relations between mathematical statistical results of the theory of continuous fractions and the real world, I describe in this article some further mathematical details of statistics of continuous fractions, including their applications to the study of chaotic dynamical systems different from the astronomical theory of planetary systems.

4. Periodic continuous fractions

The sequence $[a_0, a_1, a_2, \dots]$ of elements of the continuous fraction of the number

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}$$

can be an arbitrary sequence of natural numbers. A continuous fraction is called *periodic* if this sequence is periodic (at least starting at some position): $a_{n+T} = a_n$ for all sufficiently large n. By T I will denote the length of the shortest period.

Example. The continuous fraction of the number $\sqrt{2}$ is

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cdots}} = [1, 2, 2, 2, \dots].$$

I will denote this by $[1 + [2]]$: it has period $[2]$ of length $T = 1$.

Truncating this periodic continuous fraction, we obtain very good rational approximations of the number $\sqrt{2}$ (with small denominators); for example,

$$\sqrt{2} \approx 1, \qquad \sqrt{2} \approx 1\frac{2}{5}, \qquad \sqrt{2} \approx 1\frac{12}{29} = 1.414\dots.$$

Theorem (Lagrange)**.** *A continuous fraction of an irrational real number is periodic if and only if this number is a quadratic irrationality.*

Quadratic irrationalities are (real) roots x of quadratic equations with integer coefficients $(r, p, q) \in \mathbb{Z}^3$:

$$rx^2 + px + q = 0. \tag{4.1}$$

One cannot apply the Gauss–Kuzmin statistics to these numbers directly since they form a countable set (of Lebesgue measure zero).

Computing explicitly periodic continuous fractions of a large number of quadratic irrationalities, I have noticed that the number 1 occurs approximately 40% of the time among elements of periods of all my examples, so that my statistics turned out to be similar to the Gauss–Kuzmin statistics (where $f_1 \approx 0.47$).

Therefore I have formulated (circa 1980) the following conjectures. Consider the disk $\{p^2 + q^2 \le R^2\}$ of radius R on the integer plane $\mathbb{Z}^2$ with coordinates p and q. Compute the periods of all continuous fractions of real numbers of quadratic equations

$$x^2 + px + q = 0, \tag{4.2}$$

with coefficients belonging to this disk.

Denote by $\mathcal{N}_R(k)$ the number of all equal-to-k elements of all periods for continuous fractions of the roots of equations (4.2) whose (integer) coefficients belong to the indicated disk of radius R. Consider the "frequency" of the number k,

$$\overline{N}_R(k) = \frac{\mathcal{N}_R(k)}{\sum_{m=1}^{\infty}(\mathcal{N}_R(m))}\,.$$

Conjecture 4.1. *As $R \to \infty$ the frequencies $\overline{N}_R(k)$ tend to the frequencies f_k of the Gauss–Kuzmin statistics.*

In other words, average frequencies of elements of continuous fractions of quadratic irrationalities (4.2) coincide with average frequencies of elements of continuous fractions of random real numbers x.

Of course, a similar phenomenon should be observed for the three-dimensional family of quadratic equations (4.1) (but the two-dimensional disk on the plane $\mathbb{Z}^2$ should be replaced by the three-dimensional ball in the space $\mathbb{Z}^3$).

Several years ago, returning back (swimming south) across the kilometer-wide Volga river near the Joint Institute of Nuclear Research at Dubna, I have met an unknown to me swimmer crossing the Volga to the north who told me: "About 39 year ago, in Khabarovsk, you formulated an interesting conjecture, and all these years I tried to solve it. Now, finally, I (together with my students) solved it and I came to Moscow hoping to explain the solution to you. But I was only able to meet you here, hundred kilometers north of Moscow, near Dubna, and I want to explain to you my proof of your hypothesis."

So, now the hypothesis formulated above is proved (in a series of papers by V. A. Bykovskii, M. O. Avdeeva, and others, starting with paper [10]). Despite this Far East contribution to the theory of periodic continuous fractions there still exist a large number of unanswered questions in the areas between the theory of continuous fractions, number theory, the theory of algebraic numbers on the one hand and the physics of chaotic systems, dynamical systems, mathematical statistics, ergodic theory on the other hand; these questions reveal unexpected connections between all these different theories, which are also related in a strange way (as I will explain later) to problems of informatics, logic, and the theory of algorithms.

The Gauss–Kuzmin statistics determines also the frequencies of finite combinations $(a_{n+1}, a_{n+2}, \dots, a_{n+m})$ of m consecutive elements in the expansion of a random real number in the continuous fraction.

Conjecture 4.1 and its proof discussed above can be extended to this more general case of an arbitrary m. Therefore one could expect that this statistics also describes average frequencies of such sequences of elements of continuous fractions that form complete periods (for continuous fractions corresponding to quadratic irrationalities). Here we speak about the averaging over disks of growing radius R in the plane of equations (4.2) or over three-dimensional balls in the space of equations (4.1).

However, I have already known for a long time that these expectations should not necessarily materialize, at least for the plane of equations (4.2). Elements and sequences of m elements forming parts of periods of periodic continuous fractions may satisfy the Gauss–Kuzmin statistics in the average, whereas *the statistics of complete periods* $\{a_{n+1}, \dots, a_{n+T}\}$ may be *completely different*.

This phenomenon can be explained by nonuniformity of the convergence to the Gauss–Kuzmin limits as $R \to \infty$: one may have to take into account the appearance of sequences of length T as parts of sequences of periods of continuous fractions longer than T that appear only in the discs of radius larger than R (in the space of equations (4.2)).

In fact, I knew even more: even without taking into account statistical frequencies, just on the *set-theoretic level*, many sequences $\{a_{n+1}, \dots, a_{n+T}\}$ *never appear as complete periods* of continuous fractions of roots of equations (4.2) (even though these sequences appear, with the average frequencies given by the Gauss–Kuzmin formula, as *some parts of periods* of such continuous fractions).

Example. Periodic continuous fractions of irrational roots of equations (4.2) are always palindromic: reading this (infinite) periodic sequence of elements of a continuous fraction

backward one gets the same (periodically infinitely extended) sequence. One of the simplest examples of a palindrome in English[2] is the sentence

"Don't nod".

For example, periods $[1, 1, 2]$ and $[1, 2, 1, 4]$ are palindromic because the sequence

$$\ldots, 1, 1, 2, 1, 1, 2, \ldots$$

coincides with the reverse sequence

$$\ldots, 2, 1, 1, 2, 1, 1, \ldots .$$

Similarly, in the second example, the two-sided infinite sequence

$$\ldots, 1, 2, 1, 4, 1, 2, 1, 4, 1, 2, 1, 4, \ldots$$

(without a fixed starting point) coincides with the reverse sequence

$$\ldots, 4, 1, 2, 1, 4, 1, 2, 1, 4, 1, 2, 1, \ldots .$$

I noticed the palindromicity property of periodic continuous fractions for square roots of integers many years ago. I have proved this property for the majority of such numbers in the paper [11] (published in Rio-de-Janeiro for the anniversary of the Brazilian Mathematical Society).

Later I have noticed that the same property takes place for continuous fractions of irrational roots of equations (4.2). I have formulated this property as a conjecture (since I could not prove it in all cases and two different complete proofs were obtained later by my students (F. Aicardi in Trieste and M. Pavlovskaya in San Francisco).

The palindromicity property is satisfied not only by periodic continuous fractions of the roots of equations $x^2 + px + q = 0$ and $rx^2 + q = 0$ described earlier, but for some other quadratic equations with integer coefficients

$$rx^2 + px + q = 0.$$

The question of whether it will be violated for the majority of such equations, i.e., whether equations with palindromic continuous fractions form a small part of the ball $r^2 + p^2 + q^2 \leq R^2$ of large radius R in the space of such equations, is still open; theoretically even the opposite possibility is not yet excluded and it might happen that, on the contrary, only the small part of the ball is occupied by equations for which periods of continuous fractions of roots are not palindromic.

I do not have any theorem or even a conjecture explicitly describing those finite sequences that appear as periods of continuous fractions either for square roots of rational numbers or for irrational roots of equation (4.2).

Palindromicity is, probably, just one of the (yet unknown) restrictions imposed on such "realizable" finite sequences of natural numbers (asymptotically satisfying the Gauss–Kuzmin statistics).

Passing from the set-theoretic view on the problem to the statistical one, we also come to an open problem: elements of realizable finite sequences asymptotically satisfy the Gauss–Kuzmin statistics, but asymptotic statistics of entire realizable sequences may be completely different.

The existence of a number of (yet unknown) special properties of realizable sequences follows from entropy considerations to be explained later. The number of realizable sequences of length T (with elements bounded by some constant) forms only a small part of the number of nonrealizable ones (and the palindrome property is not sufficient to explain the observed significant reduction of the number of sequences).

[2]Translator's Note: Of course, the corresponding example in the original text of the article is in Russian.

To explain these entropy considerations, consider the length $T(p,q)$ of the (shortest) period of the continuous fraction of an irrational root of equation (4.2) (the length is the same for both roots; if the roots are rational, we will set $T(p,q)=0$).

I composed tables [12] of these lengths for $p^2+q^2 \le 1000$, and these tables provided a lot of unexpected statistical facts.

Consider the disk of radius R where $p^2+q^2 \le R^2$ and average these periods $T(p,q)$ over all integral points of this disc (where the roots are real). The so-computed average arithmetic quantities $\widehat{T}(R)$ behave rather strangely, but experiments led to the following conclusions (see [12]):

Conjecture 4.2. *The average length of the period of the continuous fraction of the root of equation* (4.2) *in a ball of growing radius R grows approximately as R (i.e., proportionally to the growth of the coefficients p and q in equation* (4.2)*):*

$$\widehat{T}(R) \sim \text{const}\, R.$$

The observed values of $T(p,q)$ behave as the square root of the discriminant $\sqrt{p^2-4q}$ of the equation (4.2). Along the majority of directions in the (p,q) plane it grows as the distance R to the origin, but in the direction of the negative q axis this square root grows as $\sqrt{R}$.

Integration over the disk shows that the averages (which take into account the contribution of all directions) grow as R, thus leading to Conjecture 4.2.

The inequality $T(p,q) \le \text{const}\, R$ also looks plausible.

In the case where the discriminant vanishes (when both roots of the equation are rational), the length of the "period" is assumed to take the value $T=0$.

The shift $x \mapsto x+1$ that preserves the length of a period proves that the function T is constant along the parabolas $p^2-4q=\text{const}$. This fact leads to the conjecture that the length $T(p,q)$ of the cycle behaves as a certain function of the discriminant p^2-4q. Examples show that this function should behave like the square root function (at least approximately, when p and q are large).

It would have been interesting to analyze the behavior of the ratio

$$\mathcal{F}(p,q) = \frac{T(p,q)}{\sqrt{p^2-4q}}\,,$$

or at least of its average values over the disks $p^2+q^2 \le R^2$ and of its level lines on the (p,q) plane.

Entropy considerations provide estimates for exponential-in-T growth of the number of sequences of T natural numbers (satisfying the Gauss–Kuzmin statistics in the required case) in the form

$$\#(T) \sim Ce^{Th},$$

where h is the corresponding "entropy" (for the statistics T of independent elements k appearing with probabilities p_k, one can easily prove the classical result

$$h = -\sum_k p_k \ln p_k$$

of the combinatorics of the multinomial coefficients).

Therefore, if $\widehat{T}$ grows as R, it would have been natural to expect the growth

$$\# \sim C_1 e^{h_1 R}$$

for the number of different periods $[a_{n+1},\dots,a_{n+T}]$ of continuous fractions for roots of equations (4.2) with $p^2+q^2 \le R^2$ if these periods would have been distributed according with the Gauss–Kuzmin statistics.

At the same time, the number of different equations (4.2) with coefficients bounded by R is only of order

$$R^2 \ll C_1 e^{h_1 R}.$$

Therefore, the statistics of the period should be different from the predictions of the Gauss–Kuzmin statistics (combinations of elements of periods). However, we do not have even empirical conjectures for this statistics of periods.

My tables (in [12] for $R \le 100$) allow us also to compute the empirical average growth with R of sums of elements of periods in the disk $p^2 + q^2 \le R^2$. It seems that the sum of elements of periods of the continuous fraction for a root of equation (4.2) averaged over the disk of radius R (we denote this average value by $\widehat{\Sigma}(R)$) grows with R as a power of R, $\widetilde{\Sigma}(R) \sim R^{1+\alpha}$. This growth corresponds to the following growth of elements of a period of length $T \sim R$:

$$|a(R)| = \frac{\widehat{\Sigma}(R)}{\widehat{T}(R)} \sim R^{\alpha}.$$

The tables in [12] suggest that $|a(R)| > R^{\alpha}$ for a positive α (of order approximately $\alpha = 1/4$?). However, this conclusion is based only on experiments where $R \le 100$.

In fact, even the logarithmic in R growth of the average size of elements of periods $|a(R)|$ is sufficient for the above arguments based on the statistics of the groups of elements of periods of the continuous fraction for the root of a quadratic equation with growing coefficients. The average value of an element computed on the basis of the Gauss–Kuzmin statistics equals ∞: since the frequency f_k of the element k decreases as k^{-2} when $k \to \infty$, we obtain the (slowly) diverging expectation

$$\hat{k} = \sum_{k=1}^{\infty} k f_k = \infty.$$

For the sum of T, satisfying the Gauss–Kuzmin statistics has the expectation

$$1(Tf_1) + 2(Tf_2) + \cdots + \hat{a}(Tf_{\hat{a}}) \sim T \sum_{a=1}^{\hat{a}} \left(\frac{k}{k^2}\right) \sim T \ln \hat{a},$$

where the average number of summands $\hat{a}$ satisfies the condition $Tf_{\hat{a}} \sim 1$, i.e., $\hat{a} \sim \sqrt{T}$.

The resulting average growth of the sum of elements of the period,

$$\Sigma \sim C_2 R \ln R \qquad (\text{if } T \sim R),$$

yields the logarithmic growth of the average value of elements

$$\hat{a} > C_3 \ln R,$$

showing that for large coefficients p and q of equation (4.2), the averaged-over-the-period element of the continuous fraction for the root (in the average over the disk of radius R) grows with R at least logarithmically.

Experimental data (for $R \sim 100$) show an even faster (polynomial?) growth of $\hat{a} \sim R^{\alpha}$.

Computing the number of combinations of T independent natural elements each not exceeding $\hat{a}$, we obtain

$$\# \sim \hat{a}^T \sim R^{\beta R} \qquad (\text{for } T \sim R,\ \hat{a} \sim R^{\alpha})$$

or, at least

$$\hat{a}(T) \sim (\ln R)^R \sim e^{R \ln \ln R} \qquad (\text{for } \hat{a} \sim \ln R).$$

In both cases the growth is much faster than R^2. These estimates show that the Gauss–Kuzmin statistics cannot be applied to sequences forming complete periods (even if it holds for elements of the period and for the frequency of each fixed combination of elements of the period).

The sequences $(a_{n+1}, \dots, a_{n+T})$ forming periods of continuous fractions for roots of the majority of quadratic equations (4.2) possess some statistics (so far unknown) which do not follow from the Gauss–Kuzmin statistics for elements of these sequences.

To find these properties (generalizing palindromicity) is an excellent problem in the arithmetic of real numbers and in the arithmetic of the corresponding continuous fractions.

It would be interesting to analyze whether there exist some restrictions on the statistics of periods of general quadratic equations (4.1) (with an arbitrary leading coefficient r); all my experimental results are for the case $r = 1$ of equations (4.2) and do not reveal any deviation of the statistics of periods for more general equations (4.1) from the Gauss–Kuzmin statistics in the case $r > 1$.

On the set-theoretic level the property to form the period of the continuous fraction for a root of a quadratic equation (4.1) does not pose any restrictions: by the (inverse) Lagrange theorem each finite sequence of natural numbers can be realized as the period of the continuous fraction for the root of equation (4.1).

However, this *set-theoretic* fact does not say anything about the *statistics* of periods: it can differ significantly from the Gauss–Kuzmin statistics providing the frequency of this sequence as a part of the continuous fraction of a random real number.

This discussion of the asymptotic behavior of various averages is far from being mathematically rigorous. For example, the asymptotic description of average values $\hat{a}$ and $\hat{b}$ of variables a and b does not yield a mathematically rigorous conclusion about the average value of a/b:

$$\widehat{a/b} \neq \frac{\hat{a}}{\hat{b}}.$$

Ignoring this obvious problem, I often follow not the mathematical point of view, but the point of view of natural sciences and freely replace the left-hand side of the previous inequality with the right-hand side (in the hope that even though this is not justified mathematically, the ratio of average values in the right-hand side provides some reasonable information that often allows us to learn something about the average value of the ratio a/b and to postpone the discussion of the difference between the ratio of averages and the average of ratios).

Example. Denote by $\tau(n)$ the number of natural divisors of the integer n:

n	1	2	3	4	5	6	7	8	9	10	11	12
τ	1	2	2	3	2	4	2	4	3	4	2	6

The average behavior of this (chaotically oscillating) function for large n is described by the following theorem (likely due to Dirichlet, but also proved by many others):

$$\left(\hat{\tau}(n) := \frac{1}{n}\sum_{m=1}^{n} \tau(m)\right) \sim \ln n.$$

(Equivalence $\sim$ means that $\lim_{n\to\infty}(\hat{\tau}(n)/(\ln n)) = 1$.)

The average behavior of the (also chaotically oscillating) sum $\sigma(n)$ of all $\tau(n)$ divisors of the number n:

n	1	2	4	5	6	7	8	9	10	11	12
σ	1	3	7	6	12	8	15	13	18	12	28

for large n is described by the following theorem (likely due to Dirichlet as well, but also proved earlier) yielding the Cesaro means

$$\hat{\sigma}(n) \sim cn,$$

where the constant $c \approx 3/2$ equals the value of the zeta function

$$\zeta(2) = \sum_{m=1}^{\infty} m^{-2} = \frac{\pi^2}{6}$$

(for the sum of squares of divisors we would have obtained the Cesaro means $\hat{\sigma}_2(n) \sim \zeta(3)n^2$, and for the sum of r-th powers the Cesaro means $\hat{\sigma}_r(n) \sim \zeta(n+1)n^r$).

Now we consider all $\tau(n)$ divisors of an integer n (for example, divisors $1, 2, 3, 6$ of the number $n = 6$). Denote by $d(n)$ the *average* of all these divisors:

$$d(n) = \frac{\sigma(n)}{\tau(n)}\,.$$

This function of n also oscillates rather chaotically. To understand the asymptotic behavior of $d(n)$ for large n, consider the Cesaro mean

$$\hat{d}(n) := \frac{1}{n}(d(1) + d(2) + \cdots + d(n)).$$

Knowing the asymptotic behavior of numerators and denominators,

$$\hat{\sigma}(n) \sim cn, \qquad \hat{\tau}(n) \sim \ln n,$$

it is natural to assume that for $n \to \infty$ the average value of the ratio behaves as the ratio of average values of the numerator and denominator,

$$\hat{d}(n) \sim \frac{cn}{\ln n}\,.$$

However, this naive "asymptotic formula" is wrong and the correct asymptotics is much larger:

$$\hat{d}(n) \sim \frac{c'n}{\sqrt{\ln n}} \gg \frac{\hat{\sigma}(n)}{\hat{\tau}(n)}\,.$$

In this problem the correct answer is known, but in a number of other similar problems, correct mathematical statements have not been proved or even formulated.

Example. Consider the sequence consisting of remainders of division by an integer n of terms of a geometric progression,

$$\{a^x \pmod{n}\} \qquad (x = 1, 2, \dots),$$

with base a relatively prime with n.

Euler proved that this sequence of remainders is periodic with period T, which is a divisor of the value of the following Euler function at the point n:

$$\hat{\varphi}(n) = (\text{the number of divisors of } n \text{ relatively prime with } n).$$

Initial values of this Euler function also oscillate quite chaotically:

n	2	3	4	5	6	7	8	9	10	11	12	p^a
φ	1	2	2	4	2	6	4	6	4	10	4	$(p-1)p^{a-1}$

The averaged *asymptotic of the Euler function* was computed by Euler himself:

$$\hat{\varphi}(n) \sim c_1 n, \qquad c_1 = \frac{1}{\zeta(2)} \approx 0.6,$$

where $\hat{\varphi}$ is the Cesaro mean:

$$\hat{\varphi}(n) := \frac{1}{n}(\varphi(1) + \varphi(2) + \cdots + \varphi(n)).$$

The proof of this asymptotic formula (using straightforward probabilistic arguments) led Euler to the discovery of the surprising multiplicative formula for his zeta function

$$\zeta(m) := \sum_{n=1}^{\infty} \frac{1}{n^m}\,,$$

which related this infinite sum over integers n with the infinite product over primes p:

$$\zeta(m) = \prod_{p=2}^{\infty} \frac{1}{1-\dfrac{1}{p^m}}.$$

The "physical meaning" of the Euler formula consists in the *computation* of the probability $1/\zeta(2)$ of the following event: integers u and v are relatively prime, so that the fraction u/v is reduced.

Indeed, the probability that an integer vector with components (u, v) is divisible by p is $1/p^2$, so that $1 - 1/p^2$ is the probability that the vector is not divisible by p. Taking the product over all prime p, we get the probability $\prod(1 - 1/p^m)$ that all m components of a vector in the space $\mathbb{Z}^m$ are not divisible by a common integer.

Substituting instead of the value $\varphi(n)$ of the Euler function its average value $c_1 n$ and replacing an unknown divisor T of the integer $N = \varphi(n)$ with the value at N of the average divisor $\hat{d}(N)$ $(\sim cN/\sqrt{\ln N})$, we would have obtained for the period T the (both unproved and wrong) average value

$$\widehat{T}_?(n) \sim \hat{d}(\varphi(n)) = \frac{c' c_1 n}{\sqrt{\ln(c_1 n)}}\,.$$

Experimental study of the actual periods $T(n)$ in the Fermat–Euler geometric progressions performed by F. Aicardi (see [13]) for n up to values of order 10^{10} indicates a different behavior of Cesaro means:

$$\widehat{T}(n) \sim \frac{c_2 n}{\ln n}\,.$$

The difference may be caused by the fact that the special divisor $N = \varphi(n)$ can be systematically different from the averaged divisor $\hat{d}(N)$.

Moreover, it may happen that the average behavior of the divisors of numbers $\mathcal{N}$ of the form $\mathcal{N} = \varphi(n)$ systematically differs from the behavior of the divisors of the majority of integers n (both for τ, for σ, and for d). Indeed, arithmetic properties of numbers

$$n = \prod_p p^{a(p)} \quad \text{and} \quad \varphi(n) = \prod_p (p^{a(p)-1}(p-1))$$

may have different statistical characteristics.

Which of the possibilities described above cause the replacement of $\sqrt{\ln n}$ in the erroneous asymptotics $\widehat{T}_?(n)$ with $\ln n$ in the experimental Cesaro means $\widehat{T}(n)$ is a very interesting question, both from the point of view of fundamental mathematics (proofs of theorems) and from the point of view of numerical experiments (where empirical average values of numbers $d(\varphi(n))$ should be compared with the behavior of numbers $\hat{d}(\hat{\varphi}(n))$, and the behavior of the periods $T(n)$ with the behavior of the function $d(\varphi(n))$).

Empirical results often differ from mathematically rigorous statements because such "growing-to-∞" functions as $\ln\ln n$ remain practically constant at any reasonable range of values of n.

For example, even for a huge integer n with 1000 digits in the 10-ary representation, the value of $\ln\ln n$ does not exceed 3.

Therefore, the function $\ln\ln n$ is empirically bounded by a "constant" C, whereas for a mathematician this function is not asymptotically equivalent to a constant since it tends to infinity as $n \to \infty$.

The *erroneous* "asymptotics" $\ln\ln n \sim \text{const}$ is *more useful* in many practical applications than the *mathematically correct but practically misleading statement* that

$$\ln\ln n \to \infty \qquad \text{as } n \to \infty.$$

These nonrigorous "physical" arguments are often more useful in many difficult problems of fundamental mathematics providing at least conjectural answers, which are either refuted or confirmed by later investigations.

Example. Studying the topological classification of smooth Morse functions (in general position) $f\colon S^2 \to \mathbb{R}$ on the two-dimensional sphere I came to the conjecture that the number $\#(T)$ of topologically distinct types of such functions with T saddle critical points (and $2T+2$ critical points overall) grows with T as T^{2T}.

This conjecture (not proved rigorously so far) was based on several (unproved) ergodic conjectures in the theory of random graphs. Rigorously I proved the inequalities

$$aT^T \le \#(T) \le bT^{2T}$$

and formulated (in [14] and [23]) the asymptotics $\#(T) \sim CT^{2T}$ as a conjecture.

The statements in the ergodic theory of random graphs on which my conjecture was based remain unproved even now. However, the conjecture about the growth of the number of types # as T^{2T} is now rigorously proved by L. Nicolaescu [15] who appended my considerations from article [14], which he had found on the Web with some ideas of quantum field theory that relate the behavior of the number $\#(T)$ with asymptotics of certain abelian integrals (imitating A. B. Givental's proof of the "mirror symmetry" phenomenon in quantum field theory).

The pseudorandomness theories discussed in this article provide numerous (although usually nonrigorous) methods for the analysis of many difficult problems of fundamental mathematics. I will discuss several more examples of such applications of physics to mathematics below.

The converse application of mathematical results to numerous problems in physics is even more obvious, and I leave to the reader the pleasure of adding new such applications to those described in this article.

5. Continuous fractions of eigenvalues and the direction of eigenoscillations

The quadratic equation

$$x^2 + px + q = 0 \tag{5.1}$$

is the characteristic equation of the second order matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad p = -\operatorname{tr} A, \qquad q = \det A.$$

For matrices with integral elements

$$A \in \operatorname{End}(\mathbb{Z}^2), \qquad A \in \operatorname{SL}(2,\mathbb{Z})$$

the coefficients p and q are integers and we can use the theory of continuous fractions for the roots of equation (5.1) discussed in Section 4.

Consider, for example, the ball B_M of radius $\sqrt{M}$ in the space of matrices:

$$\{a^2 + b^2 + c^2 + d^2 \le M\}.$$

Each point of this ball for which $p^2 - 4q > 0$ determines the characteristic equation (5.1) whose roots are quadratic-irrational (or, sometimes, rational) real numbers with, therefore, periodic continuous fractions.

The statistics of various elements of periods of these continuous fractions for all matrices (e.g., frequencies of elements of these periods) and their asymptotic behavior as $R \to \infty$ could have been deduced from results of Section 4. However, the study of these statistics (initiated in [16]) does not reduce to the application of the Gauss–Kuzmin statistics for the following reason.

Consider the map (Viète-type)

$$V\colon B(M) \to \mathbb{Z}^2$$

that associates to each matrix A its characteristic equation

$$V(A) = (p = -\operatorname{tr} A, \qquad q = \det A).$$

For each point $(p, q) \in \mathbb{Z}^2$, its full preimage in $B(M)$ consists of several matrices, and the number $|V^{-1}(p, q)|$ of these matrices depends on the point (p, q). Therefore the averaging (over the disk $\{p^2 + q^2 \le R^2\}$ in the plane $\mathbb{Z}^2$ of equations (5.1)) should be made with weights

$$w(p, q) = \frac{|V^{-1}(p, q)|}{|B(M)|}$$

that take into account the number of appearances of different points (p, q) (i.e., different traces and determinants) for matrices A in the ball $B(M)$ (whereas in taking the average in Section 4 each point of the disk $\{p^2 + q^2 \le R^2\}$ had the same weight).

The computation of the asymptotics of weights w and of the corresponding weighted asymptotics of various events is not that simple, and here I will only indicate (following [16]) some results of these computations that modify the Gauss–Kuzmin statistics in two different ways: weights and modifications turn out to be different for the space of all 2×2 integral matrices $A \in \operatorname{End}(\mathbb{Z}^2)$ and for the group of unimodular matrices,

$$A \in \mathrm{SL}(2, \mathbb{Z}) \qquad (\text{with } \det A = 1).$$

Statistics of continuous fractions of eigenvalues for integral matrices of the second order. For a generic integral 2×2 matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, the coefficients of the characteristic equation

$$x^2 + px + q = 0$$

are $p = -(a + d)$, $q = ad - bc$, and the discriminant of the equation takes the form

$$\Delta = p^2 - 4q = (a - d)^2 + 4bc.$$

On the sphere

$$a^2 + b^2 + c^2 + d^2 = N \tag{5.2}$$

(of radius $r = \sqrt{N}$) the function $|p|$ is (in the average) of order $\sqrt{N}$, whereas $|q|$ and $|\Delta|$ are of order N.

Therefore, $\sqrt{\Delta}$, as well as $|p|$, are of order $\sqrt{N}$ (in the region $\Delta \ge 0$).

According to Section 4, the average length of the period of the continuous fraction for a real root of our quadratic equation on the sphere (5.2) reaches the value $\widetilde{T}(N) \sim \sqrt{\Delta} \sim \sqrt{N}$.

To obtain a similar behavior of the average value of the period length in the ball

$$a^2 + b^2 + c^2 + d^2 \le N, \tag{5.3}$$

one has to compute weights of different spheres making up the ball (5.3). (These weights are proportional to the number of integral points on these spheres.)

The four-dimensional ball (5.3) of radius $\sqrt{N}$ contains approximately $c(\sqrt{N})^4 = cN^2$ integral points. These points lie on N spheres of various radii (whose squares are integers not exceeding N). Arguing naively, we can expect that the largest of these N spheres contains "in the average" approximately $c'N$ integral points.

The same average value can be obtained using the following more rigorous arguments. Let us subtract from the number of integral points in the ball (5.3) the number of integral points in the next closest ball that has integral points (its radius equals $\sqrt{N-1}$). The difference

$$cN^2 - c(N-1)^2 = 2cN - c \sim c'N$$

is the required (averaged) asymptotics.

Taking into account the computed weights of different spheres and considering all M spheres (5.2) with $N \le M$, we obtain the following expression for the sum of period lengths of continuous fractions of real eigenvalues for all matrices A in the ball of radius $\sqrt{M}$:

$$\sum_{N \le M} (\widetilde{T}(N)c'N) \sim \sum_{N \le M} (c''N^{3/2})$$

(since the average period length $\widetilde{T}(N)$ grows as $\sqrt{N}$).

Replacing the latter sum with an integral we obtain the value (approximately) $c'''N^{5/2}$ for the sum of approximately c_4N^2 terms (corresponding to the matrices A in our ball that have real eigenvalues).

The obtained approximate average period length for continuous fractions of real eigenvalues of integral matrices $A \in \mathrm{End}(\mathbb{Z}^2)$ in the ball (5.3) of radius $\sqrt{M}$ takes, according to [16], the value

$$\widehat{T}(M) = (c'''M^{5/2})/(c_4M^2) = c_5\sqrt{M}, \tag{5.4}$$

which confirms the similar naive (computed above) asymptotics for the value of this period length averaged over the boundary of this ball.

Mathematical tradition requires viewing this asymptotics (5.4) as a conjecture (although from the point of view of natural sciences the above computations and the article [16] give a sufficient justification of this conjecture).

Statistics of continuous fractions for eigenvalues of unimodular matrices of the second order. Passing to the group $\mathrm{SL}(2,\mathbb{Z})$ of unimodular matrices of the second order (where $\det A = 1$), we encounter a quite different behavior of continuous fractions of eigenvalues.

Theorem (see [16])**.** *The average value $\widehat{T}(N)$ of the period length of a continuous fraction for a real eigenvalue of a matrix $A \in \mathrm{SL}(2,\mathbb{Z})$ (averaging is over the ball* (5.3)*) tends to* 2 *as the radius of the ball tends to infinity:*

$$\lim_{N\to\infty} \widehat{T}(N) = 2.$$

To prove this theorem let us compute continuous fractions of the roots of characteristic equations

$$x^2 + px + 1 = 0 \tag{5.5}$$

($q = 1$ since $\det A = 1$ for $A \in \mathrm{SL}(2,\mathbb{Z})$).

These roots are real if (and only if) $|p| \ge 2$, and below we will consider this case.

Lemma 5.1. *The period length $T(p)$ of the continuous fraction of a root of equation* (5.5) *with integral coefficient p equals*

$$T(\pm 2) = 0, \qquad T(\pm 3) = 1, \qquad T(p) = 2 \text{ for } |p| \ge 4.$$

Proof. Let us consider the case $p > 0$ (the case $p < 0$ can be considered similarly or can be reduced to the case $p > 0$ by changing the sign of x).

Consider the 2-periodic continuous fraction with period $[1, a]$:

$$t = \cfrac{1}{1 + \cfrac{1}{a + t}}.$$

The 2-periodicity condition yields a quadratic equation for t:

$$1 + \frac{1}{a+t} = \frac{(a+1)+t}{a+t}, \qquad t = \frac{a+t}{(a+1)+t},$$
$$t^2 + (a+1)t = a + t, \qquad t^2 + at - a = 0.$$

Substituting $t = x + 1$ in the latter equation, we obtain the following equation for x:

$$x^2 + 2x + 1 + ax + a - a = 0, \qquad x^2 + (a+2)x + 1 = 0.$$

Taking $a = p - 2$ we arrive at (5.5).

Therefore, the root $x = t - 1$ of equation (5.5) has the following continuous fraction:

$$x = -1 + t = [-1 + [1, a]].$$

For the second root of the quadratic equation (5.5) the computations are similar.

In the special case $p = 3$ we have $a = 1$, so that the period length equals 1 (and the root yields the golden ratio $[1, 1, \dots]$). In the special case $p = 2$ both roots of (5.5), $x_{1,2} = -1$, are rational; hence $T = 0$.

Lemma 5.1 is proved. □

To deduce the theorem from Lemma 5.1 we compute weights of different values of the trace (i.e., different values of p) in the ball (5.3) in the group $\mathrm{SL}(2, \mathbb{Z})$.

Lemma 5.2. *The number of matrices $A \in \mathrm{SL}(2, \mathbb{Z})$ with $\operatorname{tr} A = s$ such that the sum of squares of all four entries does not exceed M equals*

$$Q_s(M) \le C_s M^{5/6}$$

(for some constant C_s independent of M).

Proof. For the matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ our conditions read

$$a + d = s, \qquad \det A = 1, \qquad a^2 + b^2 + c^2 + d^2 \le M.$$

Therefore, $|a| \le \sqrt{M}$, $|d| \le \sqrt{M}$, and for a fixed a the number $d = s - a$ is completely determined, whereas the point (b, c) lies on the hyperbola $bc = P$ (where $P = s(a - s) - 1$ since $\det A = 1$).

The number I of integral points on this hyperbola satisfying the condition $b^2 + c^2 \le M$ does not exceed $C\sqrt[3]{M}$ (where C is a constant independent of M and of P).

Indeed, integral points on each of two connected branches of the hyperbola in the disk $b^2 + c^2 \le M$ are vertices of a polyhedron (with integral vertices) with the area S not exceeding $C_1 M$. The number of vertices of an area S convex polyhedron with integral vertices does not exceed $C_2\sqrt[3]{S}$ (as it was proved in [26]).

Therefore, the required number I of integral points on both branches of the hyperbola is bounded from above by the number $2C_2\sqrt[3]{S}$, so that

$$I \le 2C_2\sqrt[3]{C_1 M} = C\sqrt[3]{M}.$$

Therefore, the total number $Q_s(M)$ of points obtained for all values of a does not exceed

$$(\#a)I \le (2\sqrt{M} + 1)C\sqrt[3]{M} \le C_s M^{5/6},$$

and Lemma 5.2 is proved. □

Lemma 5.3. *The total number $G(M)$ of all matrices $A \in \mathrm{SL}(2,\mathbb{Z})$ with real eigenvalues such that the sum of the squares of its four entries does not exceed M satisfies $G(M) \geq LM$, where $L > 0$ is a constant independent of M.*

Proof. The assumptions of Lemma 5.3 put the following restrictions on the matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$:

$$bc = ad - 1, \qquad a^2 + b^2 + c^2 + d^2 \leq M, \qquad (a+d)^2 \geq 4.$$

The vector (a, b) lies in the disk of radius $\sqrt{M}$. The number of integral points in this disk grows with M as the area πM of the disk. The probability that the integers a and b have a nontrivial common divisor equals $1/\zeta(2) = 6/\pi^2$ (by the Euler theorem; see pp. 49–50).

Therefore, the number of such primitive integral vectors in the disk of radius $\sqrt{M}$ grows with M as $(6\pi/\pi^2)M$. In particular, it is greater than M.

For a fixed (primitive) integral vector (a, b) the condition $bc = ad - 1$ means that the point (c, d) is on the affine straight line l parallel to the line $\mathbb{R}(a, b)$ and containing integral points.

The distance between two neighboring integral points on this line equals $\sqrt{a^2 + b^2}$. The disk $\{c^2 + d^2 \leq M - (a^2 + b^2)\}$ on the plane with coordinates (c, d) has the radius $\sqrt{M - r^2}$ (where $r^2 = a^2 + b^2$). The number of integral points on the line l inside this disk is approximately $2\sqrt{M - r^2}/r$ (because the line l is not far from a diameter of the disk).

Therefore, the total number of all matrices A is approximately given by the following sum:

$$\mathcal{F}(M) = \sum_{(a,b)} \frac{2\sqrt{M - r^2}}{r},$$

where the summation is over all primitive integral vectors in the disk $\{a^2 + b^2 \leq M\}$.

Replacing the sum $\mathcal{F}$ with the corresponding integral over this disk on the Euclidean plane with Cartesian coordinates (a, b) and taking into account the expression for the area in polar coordinates,

$$da \wedge db = r dr \wedge d\varphi,$$

and the formula for the Euler density $6/\pi^2$ of the set of primitive integral points in $\mathbb{Z}^2$ (see Section 4, p. 50), we obtain the following approximate expression for $\mathcal{F}$:

$$\mathcal{F}(M) \approx 2\pi \int_0^{\sqrt{M}} \left(\frac{6}{\pi^2}\right) 2\sqrt{M - r^2}\, dr = 6M,$$

which proves Lemma 5.3 (the additional condition that the eigenvalues be real, $p^2 \geq 4$, can be easily taken into account; it will just change the coefficient 6).

Combining Lemma 5.2 with Lemma 5.3, we find that as $M \to \infty$, the weights of events $|p| = 2$ and $|p| = 3$ (for matrices $A \in \mathrm{SL}(2,\mathbb{Z})$ with real eigenvalues and the sum of squares of all four entries not exceeding M) tend to zero as $M \to \infty$:

$$\lim_{M\to\infty} \frac{2(c_2 + c_3)M^{5/6}}{LM} = 0.$$

Therefore, the weight of the set of matrices such that $|p| < 4$ tends to 0 as $M \to \infty$.

Now Lemma 5.1 implies that the average value $\widehat{T}(M)$ of the period length of a continuous fraction tends to 2 as $M \to \infty$. Indeed, the set of matrices A with $|p| = 2$ and with $|p| = 3$ (where $T(A) = 0$ and $T(A) = 1$, respectively) becomes nonessential in the

asymptotics of the average as $M \to \infty$, whereas the matrices A with larger $|p|$ and larger trace (for which $T(A) = 2$) provide exactly this average,

$$\lim_{M\to\infty} \widehat{T}(M) = 2.$$

The theorem on page 53 is therefore proved (up to some small gaps in the proof that will be filled in later). □

Corollary. *Continuous fractions of eigenvalues of matrices $A \in \mathrm{SL}(2,\mathbb{Z})$ do not satisfy the Gauss–Kuzmin statistics for elements of continuous fractions of random real numbers.*

This means that the number $\nu_k(M)$ of times k is found as an element of the period of the continuous fractions for an eigenvalue of a matrix $A \in \mathrm{SL}(2,\mathbb{Z})$ with $a^2+b^2+c^2+d^2 \le M$ divided by the total number of all elements of all these periods does not tend, as $M \to \infty$, to the Gauss–Kuzmin frequency f_k at least for some k:

$$\lim_{M\to\infty} \frac{\nu_k(M)}{\tau(M)} \ne f_k. \tag{5.6}$$

Proof. Consider the case $k = 1$ where the frequency f_k is maximal:

$$f_1 = \frac{1}{\ln 2} \ln\left(1 + \frac{1}{k(k+2)}\right) = \log_2 \frac{4}{3} \approx 0.47 < \frac{1}{2}$$

(this frequency is less than $1/2$ because $\sqrt{2} > 4/3$).

Periods of continuous fractions of irrational eigenvalues of our matrices A all have the form $[1]$ or $[1, a]$.

Therefore, the left-hand side of the limit formula (5.6) is precisely $1/2$ (even for $\nu_k(M) > (1/2)\tau(M)$) due to the contribution of the cases $p = \pm 3$ where the period is $[1]$).

These contributions do not affect the limit as $M \to \infty$ (the same is true for matrices in $\mathrm{SL}(2,\mathbb{Z})$ with trace ± 2 for which $T = 0$) since these weights asymptotically vanish as $M \to \infty$ (by Lemmas 5.2 and 5.3). □

Remark. The bound $M^{5/6}$ for the number $Q_s(M)$ in Lemma 5.2 looks unrealistically high. A much stronger (likely realistic) estimate is given by the inequality

$$Q_2(M) \le K\sqrt{M} \ln M, \tag{5.7}$$

proved in [16] for matrices with traces 2 and -2. It is likely that the number $Q_s(M)$ of matrices with other traces behaves similarly for large M.

Example. Article [16] contains the empirical values

M	2	3	6	11	18	27	38	51	66	70
Q_2	1	5	13	17	25	37	45	49	53	53

yielding, for example, the values

M	11	38	66
$\frac{Q_2}{\sqrt{M}}$	5.04	7.30	6.52

It is not even clear whether we need the growing factor $\ln M$ in inequality (5.7) or if the estimate of the form

$$Q_2(M) \le K'\sqrt{M}$$

holds.

The above estimates for average period length, which are not rigorously proved (but are quite reliable) depend on the following "weakly asymptotic" ergodic theorems that reduce the analysis of sums to the estimation of integrals.

Consider the subset M of the lattice $\mathbb{Z}^m$ consisting of all "primitive" integral vectors $(x_1, \dots, x_m)$, i.e., such vectors that their components do not have a common divisor greater than 1. The "ergodic" statement of the theory of weak asymptotics claims that this set M of primitive vectors, which has average density $1/\zeta(m)$ (as explained earlier on pp. 49–50) is *asymptotically uniformly distributed in* $\mathbb{Z}^m$.

The set M is by no means dense: there are arbitrarily large holes in $\mathbb{Z}^m$ that are free of primitive vectors.

An *asymptotically uniform distribution* is defined as follows.

Choose a "test cone K" with vertex at the origin of the Euclidian space $\mathbb{R}^n$ (it suffices to consider a polyhedral cone bounded by a finite set of hyperplanes passing through the origin).

Consider the part $M \cap K$ of M lying in K. Denote by $\#_R(K)$ the number of points x in $M \cap K$ inside the ball $\|x\| \le R$ of radius R.

Denote by $V_R(K)$ the total number of all integral points in the cone K inside the same ball of radius R.

The set M is called *asymptotically uniformly distributed in* $\mathbb{Z}^m$ if the "frequency" $f_K(M)$ that a point in M hits the cone K, defined as the limit

$$f_K(M) = \lim_{R\to\infty} \frac{\#_R(K)}{V_R(K)},$$

exists and does not depend of the "test cone" K.

In our example where M is the set of all primitive vectors this limit frequency of M should be the constant

$$\rho = f_k(M) = \frac{1}{\zeta(m)},$$

representing the "probability that a point $x \in \mathbb{Z}^m$ is primitive".

Corollary. *Let $f\colon \mathbb{R}^m \to \mathbb{R}$ be a smooth function vanishing outside of a ball B of finite radius. Choose a small "Planck constant" h and compute the average value (over the points x of the set M)*

$$I_h(f) = \Bigg(\sum_{x\in M} f(hx)\Bigg)\Big/(Z(h)),$$

where $Z(h)$ is the number of points hx, $x \in M$, inside the ball B.

Then the fact that M is uniformly distributed is equivalent to the relation

$$\lim_{h\to 0} I_h(f) = \frac{\int_B f(z)\,dz}{\int_B 1\,dz}$$

for each smooth function f (vanishing outside the ball B).

This result yields "weak asymptotics" of the function f on the set M:

$$\sum_{\substack{x\in M\\ \|x\|\le R}} f(x) \sim \rho \int_{B(R)} f(z)\,dz$$

for many "reasonable" functions f.

Here the notation "$\sim$" means that the ratio of the left-hand side to the right-hand side tends to 1 as the radius R of the ball $B(R)$ tends to infinity.

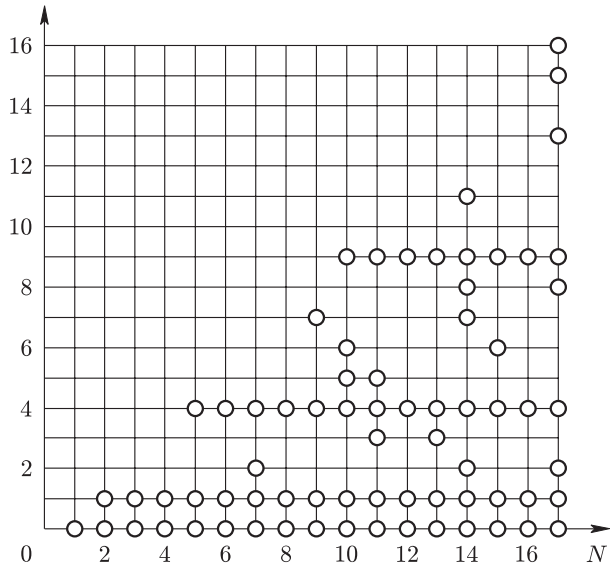

FIGURE 1

The class of "reasonable" functions includes, in particular, all harmonics

$$f(z) = e^{i\langle k,z\rangle}$$

(hence all trigonometric polynomials).

We used the above relation in the case of the function $f = \sqrt{M - r^2}$, which is also "reasonable".

It might look strange, but I have not seen a proof of the uniform distribution property either for the set of primitive integral vectors, or for numerous other asymptotically uniformly distributed sets naturally generalizing this one (which are quite common, for example, in the pseudoperiodic topology of quasicrystalls; see [21]).

One of the simplest examples of such a set is the set of critical points of a pseudoperiodic potential resembling multidimensional generalizations of the one-dimensional potential

$$f(x) = \sin x + \sin(\sqrt{2}x),$$

defined for $x \in \mathbb{R}^1$.

Another interesting example is provided by quadratic residues. The remainder x of division by an integer N is called a *quadratic residue* if there exists an integer m such that the remainder of the division of m^2 by N equals x.

The pairs (x, N), where x is a quadratic residue modulo N, form a chaotically looking subset of the integral lattice with coordinates (x, N) (see Figure 1).

The conjecture is not only that these points are distributed asymptotically uniformly with density $\rho = 1/2$ on the (x, N) plane in the sector $0 \le x < N$), but that the Kolmogorov stochasticity parameter λ_n of the set of all quadratic residues modulo N takes values close to Kolmogorov's average value $\Lambda_* \approx 0.86$ (which tend neither to zero nor to infinity as $n \to +\infty$).

Example. For $N = 20$ we have $\lambda_{10} \approx 0.79$; for $N = 19$, $\lambda_{19} \approx 0.19$. Kolmogorov's distribution Φ yields the following probabilities of smaller values of λ: $\Phi(0.79) \approx 0.456$; $\Phi(0.91) \approx 0.717$.

6. Multidimensional continuous fractions

The above theory of periodic continuous fractions for quadratic irrational numbers, as well as its applications to the statistics of continuous fractions for eigenvalues of 2×2 matrices, can be generalized to algebraic numbers of higher degree and to integral matrices of higher order.

The most natural generalization is provided by the notion of a multidimensional continuous fraction introduced by F. Klein. Even though these algebraic numbers and eigenvalues have ordinary continuous fractions, their behavior seems to be both more chaotic and less understood, whereas for Klein's multidimensional continuous fractions there is a nice general theory which generalizes, in a natural way, the periodicity of continuous fractions of quadratic irrationalities.

Let us begin with a simplicial cone $K \subset \mathbb{R}^n$ which is one of 2^n connected components of the complement in the Euclidean space $\mathbb{R}^n$ to the union of n hyperplanes (in general position) passing through the origin.

We will view the space $\mathbb{R}^n$ as containing the standard (rank n lattice $\mathbb{Z}^n \subset \mathbb{R}^n$ and a "multidimensional continuous fraction" will be a geometric object determined by the relative position of the cone K and the lattice $\mathbb{Z}^n$.

The intersection $P = \mathbb{Z}^n \cap K$ is an additive subgroup of the vector space $\mathbb{R}^n$. Its convex hull $\widehat{P}$ is an (infinite) convex polyhedron (with integral vertices) in the space $\mathbb{R}^n$. The boundary $V = \partial\widehat{P}$ of this polyhedron is an $(n-1)$-dimensional (polyhedral) hypersurface with integral vertices belonging to P.

This hypersurface V (homeomorphic to the space $\mathbb{R}^{n-1}$) is called the *sail* (of the symplicial cone K). Such sails V are n-dimensional generalizations of continuous fractions (corresponding to the case $n = 2$ and one-dimensional sails, which are convex broken lines in the plane $\mathbb{R}^2 \supset \mathbb{Z}^2$ having integral vertices and consisting of an infinite number of straight intervals).

In this special case $n = 2$ the geometry of the sail is described by the following positive integers. Each straight interval a of the broken line V has its *integral length* defined as the number of pieces to which a is divided by its integral points.

Example. Intervals joining the points $(0,0)$, $(2,0)$, and $(0,3)$ have integral length 2, 1, and 3, respectively.

To each vertex of the broken line with integral vertices we associate a positive integer called the "integral angle" of this broken line. The *integral angle* between the intervals OA_1 and OA_2 with integral endpoints at their common vertex O is defined as the area of the parallelogram with sides OA_1 and OA_2 divided by the product of the integral lengths of its sides OA_1 and OA_2.

If the plane and the angle are oriented, then the integral angle (which is defined below as an integer) can be naturally supplied with the sign,

$$\text{integral angle } (OA_1, OA_2) = \det \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = \frac{[OA_1, OA_2]}{|OA_1||OA_2|},$$

where (x_k, y_k) is the smallest integral vector on the ray OA_k; here $[\xi, \eta]$ denotes the oriented area of the parallelogram with sides xi and η.

Let us note that the (positive) integral angle is the index of the sublattice, in the lattice $\mathbb{Z}^2$, generated by the intersection of $\mathbb{Z}^2$ with the lines OA_1 and OA_2.

Example. Integral (nonoriented) angles of the triangle with vertices $(0,0)$, $(2,0)$, and $(0,3)$ are $1, 3$, and 2, respectively.

Now we assume that one of the sides of the angle K in $\mathbb{R}^2$ is the line given by the

equation $y = kx$, $k > 0$ (in the natural coordinate system where integral points form the standard lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$) and $y < kx$ in K.

In this case, consecutive integral lengths of edges of the sail numbered towards going to infinity along the indicated line ($x \to +\infty$) turn out to be equal to consecutive elements $(a_0, a_2, a_4, a_6, \dots)$ of the continuous fraction of the number

$$k = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}},$$

with even indices, whereas consecutive integral angles between these edges are equal to the sequence $(a_1, a_3, a_5, \dots)$ of elements of the same continuous fraction with odd indices.

For the sails of the complementary angle K' where $y > kx$, consecutive integral edge lengths form the sequence $(a_1, a_3, a_5, \dots)$, and consecutive integral angles between these edges form the sequence $(a_2, a_4, \dots)$.

Now let us assume that $A \in \mathrm{SL}(n, \mathbb{Z})$ is the matrix of a linear transformation $A\colon \mathbb{R}^n \to \mathbb{R}^n$ that maps the standard lattice $\mathbb{Z}^n$ onto itself and has n distinct positive eigenvalues $(\lambda_1, \dots, \lambda_n)$.

The corresponding n invariant hyperplanes (each is generated by $n-1$ eigenvectors v_k: the first omitting v_1, the second omitting v_2, ..., the last omitting v_n) divide the space $\mathbb{R}^n$ into 2^n simplicial cones each invariant under the transformation A.

Let K be one of these invariant cones. The sail generated by this cone is the *n-dimensional continuous fraction of the operator A* (in fact, by "fraction" one can understand the entire collection of 2^n interrelated sails generated by all 2^n invariant cones of the operator A).

The invariance of the cone K and of the lattice $\mathbb{Z}^n$ under the automorphism A implies that the sail of the cone K is also invariant under A. In this sense the operator A is one of the symmetries of its sail.

In the case $n = 2$ this A-invariance of the sail means nothing but the periodicity of the corresponding usual continuous fraction (for directions k_j of eigenvectors, hence also for eigenvalues λ_j).

In the general case $n > 2$ the sail has, in addition to A, other symmetries as well. Denote by H the "integral Cartan subgroup" consisting of the operators $B \in \mathrm{SL}(n, \mathbb{Z})$ commuting with the operator A (their matrices are diagonal in the basis of eigenvectors of the operator A).

Consider the subgroup $S \subset H$ consisting of those elements of the Cartan subgroup H that preserve the cone K (rather than permuting it with another cone, as does, for example, the operator -1 belonging to H for even n).

The group S is the *symmetry group* of the sail of the cone K; it contains the operator A.

In his article [17] the Japanese mathematician Tsuchihashi used a theorem of Dirichlet to prove that the symmetry group is isomorphic to $\mathbb{Z}^{n-1}$. Its topological action of the sail (which is homeomorphic to the Euclidean space $\mathbb{R}^{n-1}$) is topologically similar to the action of the standard integral lattice $\mathbb{Z}^{n-1} \subset \mathbb{R}^{n-1}$ on the ambient Euclidean space by translations by its integral vectors.

Therefore the quotient space of the hypersurface of the sail by its symmetry group A is homeomorphic to the torus T^{n-1}.

This torus inherits from the sail a subdivision into convex faces of various dimensions; we will call this subdivision the *natural triangulation* of the sail. It also inherits the subset of integral points on each face, integral lengths, integral angles, etc.

The resulting triangulation of the torus has a finite number of faces, vertices, integral points, angles, etc. This entire structure provides a multidimensional generalization of

the *period* of a periodic continuous fraction (to which this structure is reduced in the case $n = 2$).

In this particular case the quotient torus T^{n-1} is the circle S^1 divided into N pieces (if the period length $T = 2N$ is even). If the entire period of the continuous fraction is $[a_1, a_2, \dots, a_{2N}]$, then the "triangulation" consists of N intervals of integral length $(a_2, a_4, \dots, a_{2N})$ separated by vertices with integral angles $(a_1, a_3, \dots, a_{2N-1})$.

In principle, this geometry of triangulations yields a multidimensional analog of the theory of continuous fractions of quadratic irrationalities; however, in the case $n > 2$ many interesting questions remain open.

For example, in the multidimensional case the length T of the period of an ordinary continuous fraction should be replaced with something else (not yet invented). Natural candidates for the role of the generalized period length are:

1) the number of faces of the triangulation (either only $(n-1)$-dimensional, or all);
2) the number of integral points;
3) the sum of integral volumes of faces;
4) solid angles at vertices or faces of various dimensions, their sums, alternating sums, and other natural combinations.

There are also many other candidates; for example, generalizations of theorems about the behavior of average period lengths $\widehat{T}(R)$ from the quadratic to the multidimensional case (even just to the cubic case $n = 3$) could provide reasonable arguments for the choice of a multidimensional analogue of the notion of the period length T and its average $\widehat{T}$, as well as for the choice of multidimensional analogues of the sum of elements of the period Σ and its average $\widehat{\Sigma}(R)$, or for "average elements of a period" of the form $\hat{a}(R) = \widehat{\Sigma}(R)/\widehat{T}(R)$ or of the form $\widehat{a'}(R) = \widehat{\Sigma/T}$.

The multidimensional theory also has the following fundamental gap. The period $[a_{k+1}, \dots, a_{k+T}]$ of the periodic continuous fraction of an appropriate quadratic irrationality can be an arbitrary sequence of T integers.

In the n-dimensional case triangulations of the torus T^{n-1} obtained from sails of all operators A form some strange subset of the set of all subdivisions of the torus T^{n-1} into convex polyhedra.

Conjecture. *The problem of whether a particular triangulation of the torus T^{n-1} can be realized as a natural triangulation of the sail of some integral matrix $A \in \mathrm{SL}(n, \mathbb{Z})$ is algorithmically unsolvable.*

Examples of both realizable and nonrealizable triangulations exist already in the case of the two-dimensional torus ($n = 3$).

Example (E. I. Korkina, [18, 19])**.** *The simplest realizable triangulation of the two-dimensional torus consists of two triangles, three edges, and one vertex* (see Figure 2).

This triangulation is realized by the matrix of the "three-dimensional golden ratio"

$$A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The usual golden ratio is realized by a similar matrix (which generates the Fibonacci numbers)

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

O. N. Karpenkov [20] suggested two interesting algorithms, of which the first constructs the matrix $A \in \mathrm{SL}(n, \mathbb{Z})$ realizing a given triangulation of the torus T^{n-1}, and the second provides the proof that a given triangulation is nonconstructible.

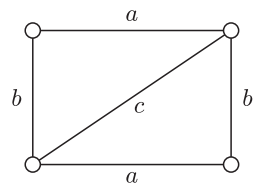

FIGURE 2

Unfortunately, for some (typical?) triangulations neither of these two algorithms ever stops: for these triangulations it is not known whether there exists an algorithm determining realizability, and I believe that such an algorithm is not only unknown, but does not exist.

The Gauss–Kuzmin statistics of elements of continuous fractions of random real numbers can also be generalized to multidimensional continuous fractions.

A dozen years ago I suggested to Yu. Sukhov to try to find such a generalization (and to find out, for example, whether the sail of a typical simplicial cone $K \subset \mathbb{R}^3$ has more triangular faces than pentagonal ones).

Instead of computing the required statistics, the resulting paper by Kontsevich and Sukhov (see [21]) proves the following axiomophilic theorem: "All similar statistical Arnold questions *have answers* (unknown to us)" and "these answers are *universal*, i.e., the resulting statistics do not depend on K (are the same) for almost all simplicial cones $K \subset \mathbb{R}^n$, so that exceptional cones where the statistics is different, form a set of Lebesgue measure zero."

Later Bykovskii and Karpenkov computed the required frequencies for various events on typical sails (frequencies of triangles, of long edges, etc.) by two different methods. Their answers are the same, although the proof that they coincide turned out to be not that easy. Both formulas generalizing the Gauss formulas for frequencies f_k are rather complicated and I leave to the reader the pleasure of finding them.

7. Statistics of periodic orbits of the Fibonacci cat map

One of the simplest models of a chaotic dynamical system is the so-called "Fibonacci cat map" $A\colon T^2 \to T^2$ of the two-dimensional torus to itself, which sends the point with coordinates (x, y) to the point

$$A(x, y) = (2x + y, x + y).$$

Fibonacci numbers appear as the coordinates of images of the basis vector $A^n(0,1) = ((1,1), (3,2), (8,5), \dots)$.

The numbers $x \pmod 1$ and $y \pmod 1$ are the coordinates (of angular type: "longitude" and "latitude") on the surface of the two-dimensional torus.

The chaotic nature of the ergodic transformation preserving the Lebesgue measure is displayed, for example, by its mixing property: for any (measurable) subset U ("the cat" U) of the torus its image after sufficiently large number iterations t of the map A will be distributed uniformly over the torus so that for each "test set" $V \subset T^2$ the part of the cat $A^t(U)$ in this set will be (asymptotically) proportional to the measure of this set:

$$\lim_{t\to\infty} \frac{\operatorname{meas}((A^t U) \cap V)}{\operatorname{meas}(A^t U)} = \frac{\operatorname{meas} V}{\operatorname{meas} T^2}\,.$$

The majority of orbits

$$\{A^t z, \quad z \in T^2, \quad t = 0.1, 2, \dots\}$$

are everywhere dense subets of the phase space T^2, and the number of times $\#(T)$ the orbit visits a test set V after T iterations ($t = 1, 2, \dots, T$) is asymptotically proportional to the measure of the test set:

$$\lim_{T\to\infty} \frac{\#(T)}{T} = \frac{\operatorname{mes} V}{\operatorname{mes} T^2}\,.$$

However, for some initial points a the orbit behaves differently. For example, there exist *periodic orbits* that return to the initial point at some moment τ in time and then continue to pass over τ points of the orbit,

$$z,\ Az,\ A^2z, \dots, A^{\tau-1}z \quad (\text{and then } A^\tau z = z).$$

Namely, periodic orbits are orbits of all points z with rational coordinates (and these points only). The period length of different periodic points is, in general, different.

Indeed, if the coordinates $x = u/n$ and $y = v/n$ of a point $z \in T^2$ are rational, with common denominator n, then the point Az also has rational coordinates with the same denominator n (since the determinant of the integral matrix $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ equals 1, so that the operator A maps the standard integral lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$ isomorphically to itself).

The phase space T^2 contains n^2 points whose coordinates have common denominator n. They form the finite set $\mathbb{Z}_n^2$ (where $\mathbb{Z}_n = Z/n\mathbb{Z}$ denotes the ring of n remainders of division by n) which is a finite version of the torus T^2 consisting, however, of only n^2 points. Our automorphism $A\colon T^2 \to T^2$ permutes n^2 points of the finite torus $\mathbb{Z}_n^2$.

Each permutation A of a finite set (consisting, for example, of N points) is decomposed (by the ergodic theorem) into *cycles* (pairwise disjoint ergodic components):

$$N = a_1 + a_2 + \dots + a_q,$$

where a *cycle of length* a consists of exactly a distinct points (z, Az, $A^2z, \dots, A^{a-1}z$), with $A^a z = z$.

The only invariant of the decomposition of a finite set (phase space of N points) into cycles of a permutation A of this set is the "*Young diagram*" of cycles: it determines the permutation (or decomposition into cycles) up to renumbering of points, which plays the role of changing the coordinate system; all possible renumberings form the group $S(N)$ of all $N!$ permutations of N points of the phase space.

The *Young diagram* consists of N unit squares on the plane. Cycles are ordered by decreasing length,

$$a_1 \ge a_2 \ge \dots \ge a_q.$$

The first row of the Young diagram consists of a_1 side-by-side squares, the second row consists of a_2 squares, etc., with the last (the shortest) q-th row consisting of a_q squares.

Example. The permutation of 9 digits given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 1 & 7 & 4 & 9 & 8 & 6 & 3 & 5 \end{pmatrix}$$

has the Young diagram of the form shown in Figure 3 because this permutation consists of $q = 4$ cycles $(3, 7, 6, 8)$, $(1, 2)$, $(5, 9)$, (4) with lengths $a_1 = 4$, $a_2 = 2$, $a_3 = 2$, $a_4 = 1$, respectively.

Keeping in mind that the dynamics of the Fibonacci cat map is chaotic, one could have thought that the induced permutations of n^2 points of finite tori $\mathbb{Z}_n^2$ would have Young diagrams (which describe the statistics of periodic motion of our dynamical system) similar to Young diagrams of cycles of random permutations of the same number $N = n^2$ of points.

Young diagrams of cycles of different permutations $P \in S(N)$ are, of course, quite different. However, one can study their statistics computing average values of various

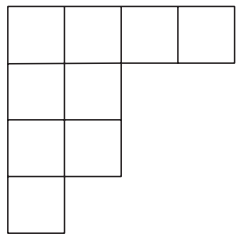

FIGURE 3

characteristics of these Young diagrams over all $N!$ permutations in the symmetric group $S(N)$.

For the simplest characteristics of Young diagrams I have chosen the following four parameters:

- *Length* l (in the previous notation $l = a_1$ is the period of the longest cycle).
- *Height* q (the number of cycles in the permutation).
- *Width* $\lambda = N/(lq)$ (this parameter shows what part of the rectangle made of q copies of the longest row of the diagram is filled with squares of the diagram).
- *Asymmetry* $\mu = q/l$ (this parameter shows whether the diagram is long, which happens when $\mu < 1$, or tall, when $\mu > 1$).

I have computed (in [22] and [23]) average values ($\hat{l}$, $\hat{q}$, $\hat{\lambda}$, $\hat{\mu}$) of these parameters for all $N!$ permutations of N elements for $N \leq 7$ (when $N! \leq 5040$). These empirical values suggest the following asymptotic behavior:

$$\hat{l} \sim c_1 N, \qquad \hat{q} \sim q_2 \ln N, \qquad \hat{\lambda} \sim \frac{c_3}{(\ln N)}, \qquad \hat{\mu} \sim c_4 \frac{\ln N}{N}\,.$$

The majority of random permutations have *long* diagrams filling only a *small part* of rectangles circumscribed around these diagrams.

Surprised by this asymmetry I studied the case of larger values of N. To avoid the summation of $N!$ summands I replaced averaging over the group $S(N)$ by the Monte-Carlo method: instead of all $N!$ permutations of, say, $N = 100$ elements (00,01, $\dots$, 99) I considered just one permutation, choosing it in a *really random* way.

As a source of random choices, in [22] and [23] I have used, first of all, *The directory of phone numbers of members of the National Academy of Sciences in Washington*. Out of the entire 7-digit number I have used only the fourth and fifth digits (to avoid the influence of geographic nonuniformity of Academy members on the first three digits).

To obtain a random permutation $s \in S(N)$ of a set with N elements one should have a family of mutually independent elements $(p_1, p_2, \dots)$ (with repetitions) of length approximately $N \ln N$.

Starting with p_1 we then take p_2 if $p_2 \neq p_1$, etc. The element p_j will occur in the resulting permutation if it is different from all previous elements $p_{<j}$.

The resulting permutation of all $N = 100$ numbers (between 00 and 99) taken from the phone directory provides a permutation $s \in S(100)$. Computation of parameters of this permutation confirmed the above conjectures about the asymptotic behavior of average values of these four parameters.

For higher reliability, I have repeated the experiment replacing phone numbers with numbers on license plates of cars passing the Steklov Mathematics Institute on the Vavilova Street in Moscow. Statistics of license plates (presented in [22] and [23]) turned out to be quite similar to the statistics obtained using phone numbers of academicians described earlier.

Finally, to augment these experiments I have considered "random permutations" of p^2 *elements of Galois fields* (from the book [3]) and taking $p = 7, 11, 13, 17, 19$.

The statistics of the resulting Young diagrams turned out to be exactly the same as for academicians and for cars, and even though my old conjecture about the (pseudo)random nature of tables of Galois fields is not yet rigorously proved, the performed experiment can be viewed as an additional argument in favor of this conjecture of fundamental mathematics.

After all these results were published, I found out that shapes of Young diagrams of random permutations were studied in the 1940s in Moscow by V. L. Goncharov [24].

Even though I (being a school student at that time) discussed with Vassily Leonidovich a lot of various mathematical problems, he never told me anything about these wonderful asymptotic results saying that

$$\hat{l}(N) \sim c_1 N, \qquad \hat{q}(N) \sim c_2 \ln N$$

(his constant c_1 was approximately 0.62).

An "explanation" of these asymptotic results can be provided by the following "nonrigorous" arguments. At repeated independent selections of one element from a box filled by many copies of each of N elements of the set we study, a repetition of a selected element can be expected to occur, on the average, after θN selections, where θ is a constant in the interval $0 < \theta < 1$ (or, at least, $\theta(N)$ tends to such a constant as $N \to \infty$).

Similar arguments show that the length of the first (the longest) cycle, even though it is different for different permutations of N elements, will have the average value $\theta' N$.

To estimate the length of the second cycle we can assume that after excluding all copies of $\theta' N$ elements in the first cycle, we obtain the action of the original permutation of N elements on the remaining $N_1 = (1 - \theta')N$ distinct elements that has the same statistics as the statistics of $N_1!$ "random" permutations forming the group $S(N_1)$.

Therefore, the expected length of the second cycle is $\theta' N_1$ and after removing all copies of elements of this cycle the box will contain $N_2 = (1 - \theta')^2 N$ distinct elements.

Similarly, after removing elements of the s-th cycle the box will contain $N_s = (1-\theta')^s N$ distinct elements, which will be permuted by the original permutation A "in a random way".

The end ($s = q$) of this process of finding cycles occurs when the number N_s of remaining distinct elements becomes small (as, for example, N_{q-1} for $a_q = 1$). This argument explains the asymptotics

$$q \sim \frac{\ln N}{-\ln(1 - \theta')}.$$

Trying to apply these results of fundamental mathematics about the statistics of random permutations to the periodic orbit of the "Fibonacci cat map" in [22] and [23], I have computed Young diagrams (of area $N = n^2$) for the corresponding automorphisms of finite tori,

$$A\colon \mathbb{Z}_n^2 \to \mathbb{Z}_n^2$$

for $n = (2, 3, \dots, 20)$. The results of these experiments turned out to be surprising. The Fibonacci cat maps permute n^2 elements in a *nonrandom way.*

I present several values of the parameters of "Fibonacci maps" of finite tori consisting of $N = n^2$ elements taken from [22] (see the table below).

The number of cycles q turns out to be much larger than the expected average value $\ln N$ for a random permutation of N elements.

The length of the longest cycle l is much less than $N = n^2$.

In [23] these empirical observations showing nonrandomness of Fibonacci permutations received a theoretical explanation. In particular, it was shown that for prime values $n > 5$

n	l	q	λ	μ
2	3	2	0.67	0.67
3	4	3	0.75	0.75
4	3	6	0.89	2.00
5	10	5	0.50	0.50
6	12	6	0.50	0.50
7	8	7	0.88	0.88
11	5	25	0.97	5.00
12	12	18	0.67	1.50
16	12	30	0.71	2.50
17	18	17	0.94	0.94
19	9	41	0.98	4.56
97	98	97	0.99	0.99

the Young diagram is almost rectangular,

$$(a_1 = \cdots = a_{q-1}; \quad a_q = 1),$$

with length either $l = n+1$ or $(n-1)/2$, and height either $q = n$ or $2n+3$, respectively.

Other (interesting) mathematical explanations from [23] based on the same observation originate from the following remark.

The transformation $A\colon \mathbb{Z}_n^2 \to \mathbb{Z}_n^2$ is not random because it is determined by a *linear* transformation A of the plane. Therefore, if we know the behavior of the orbit of a point z, we know the behavior of homothetic orbits of the points $2z$, $3z$, etc.

To eliminate the influence of this symmetry (which would have been absent in nonlinear models) we can replace the finite torus $\mathbb{Z}_n^2$ by its *projective version* (whose points are straight lines passing through the point $z = 0$).

To simplify the computations, let us assume that $n = p$ is a prime number. In this case the projectivized finite torus

$$P\mathbb{Z}_p^2 = (\mathbb{Z}_p^2 \backslash 0)/(\mathbb{Z}_p \backslash 0)$$

consists of $p+1$ points and these points are permuted by the map $A \in \mathrm{SL}(2, \mathbb{Z})$.

One could expect that the induced "*projective permutation*" of $p+1$ points would have a "more random" Young diagram than the action of A as a permutation of p^2 points of the finite torus $\mathbb{Z}_p^2$.

However, experiments show that the Young diagrams of these "projective permutations" of $p+1$ still differ significantly from Young diagrams of random permutations. For example, tables in [22] give the following values for the parameters of Young diagrams of "projective Fibonacci permutations":

n	l	q	λ	μ
2	3	1	1.00	0.33
3	2	2	1.00	1.00
5	5	2	0.60	0.40
7	4	2	1.00	0.50
11	5	4	0.60	0.80
13	7	2	1.00	0.20
17	9	2	1.00	0.22
19	9	4	0.53	0.44
41	10	6	0.70	0.60
97	40	2	1.00	0.05

Another (also presented in [23]) explanation of the special nature of Young diagrams of permutations of n^2 points of the finite torus $\mathbb{Z}_n^2$ induced by automorphisms $A \in \mathrm{SL}(2,\mathbb{Z})$ is based on the following arguments.

The permutation of $N = n^m$ points of the finite torus $\mathbb{Z}_n^m$ induced by an integer $m \times m$ matrix $A \in \mathrm{SL}(m,\mathbb{Z})$ with determinant 1 has a period $T(A,n)$ such that the matrix A satisfies the congruence

$$A^T \equiv 1 \pmod{n}.$$

Therefore all lengths of cycles of the considered permutation of points of the finite torus should be divisors of the integer $T(A,n)$.

Young diagrams of permutations for which all cycle lengths have this arithmetic property (are divisors of a fixed integer T) may have a special statistics (compared to the statistics of all permutations in $S(n^m)$).

In would be interesting to generalize Goncharov's results ($\hat{l} \sim c_1 N$, $\hat{q} \sim c_2 \ln N$) to averages over permutations in $S(n^m)$ with cycle lengths that are divisors of T and to study the dependence of this generalized statistics on the arithmetics of the number T.

Statistical properties of T (as a function of $A \in \mathrm{SL}(m,\mathbb{Z})$ and $n \in \mathbb{Z}^+$) are also interesting and nonobvious (similarly to the complexity of the behavior of periods in the Fermat and Euler theorems discussed in Section 4, pp. 49–50).

Periodic orbits of the Fibonacci cat map were studied in papers by physicists, including Dyson, Percival, Vivaldi, and others (see the survey in [25]). They computed period lengths for some special examples using nontrivial facts from algebraic number theory. However, all their examples provide only a nontypical small part of cases, which are not similar to the huge majority of other cases (both for $A \in \mathrm{SL}(m,\mathbb{Z})$ and for the original Fibonacci cat map corresponding to the automorphism $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$).

Participants of my 2005 lectures (for high school students) in the Joint Institute of Nuclear Research in Dubna (published in [23]) discovered several interesting additions to my statistical conclusions from materials of these lectures.

According to M. E. Kazaryan and V. A. Kleptsyn, in the majority of cases the Young diagrams of the projectivized Fibonacci cat map of the projective torus (consisting of $p+1$ points) are almost rectangular, of the form $l \times a$ or $(l \times a) + 1 + 1$, where an even integer a equals 2 in about 60 % of cases (these Young diagrams have $q = a$ rows in the first case and $q = a + 2$ rows in the second case).

In the table presented earlier, the first case can be seen when $p = 3, 7, 13, 17$, and the second case when $p = 11$ and 19.

According to Kazaryan and Kleptsyn, the averaged length l and height h of Young diagrams of Fibonacci automorphisms of finite tori $\mathbb{Z}_n^2$ consisting of a large number $N = n^2$ of points grow with N as $\sqrt{N} = n$, with the typical average width $\hat{\lambda} \approx 0.8$ and typical average asymmetry of order $\sqrt{N}$.

At the first glance, these three relations ($\hat{l} \sim \sqrt{N}$, $\hat{q} \sim \sqrt{\widehat{N}}$, $\hat{\mu} = \widehat{q/l} \sim \sqrt{N}$) contradict one another since $lq = \lambda N$. However, the following examples show that there is no contradiction because the average value of the ratio $\widehat{q/l}$ is not equal to the ratio of average values $\hat{q}/\hat{l}$.

Consider the population consisting of (rectangular) Young diagrams of the following three types:

$$(l,q) = (N,1) \quad \text{with probability } \alpha,$$
$$(l,q) = (1,N) \quad \text{with probability } \beta,$$
$$(l,q) = (\sqrt{N},\sqrt{N}) \quad \text{with probability } \gamma \quad (\text{here } \alpha+\beta+\gamma = 1).$$

In the case $\alpha = \beta = 1/\sqrt{N}$, $\gamma = 1 - 2/\sqrt{N}$ all three average values

$$\hat{l} = \alpha N + \beta + \gamma, \quad \hat{q} = \alpha + \beta N + \gamma, \quad \widehat{q/l} = \alpha/N + \beta N + \gamma$$

are of order $\sqrt{N}$.

Similar examples with not so rectangular diagrams allow us to reach the condition $\lambda \approx 0.8$ for the width (with $\lambda = N/(ql)$) instead of the observed in our (rectangular) example value $\lambda = 1$.

It would be interesting to find out, at least empirically, which of the discovered properties of permutations of points of a finite torus are specifically related to the Fibonacci permutation $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right)$, and which properties hold for more general cases $A \in \mathrm{SL}(2, \mathbb{Z})$ or $A \in \mathrm{SL}(m, \mathbb{Z})$.

References

[1] A. Kolmogoroff, *Sulla determinazione empirica di una legge di distribuzione,* Giorn. Ist. Ital. Attuari. **4** (1933), 83–91.

[2] ———, *On a confirmation of Mendel's laws*, Dokl. AN SSSR **27** (1940), no. 1, 38–42. (Russian) MR0003556 (2:237d)

[3] V. I. Arnold, *Dynamics, statistics, and projective geometry of Galois fields*, § 5. Adiabatic analysis of remainders of geometric progressions, MCCMO, Moscow, 2006. (Russian)

[4] ———, *Empirical study of stochasticity for deterministic chaotical dynamics of geometrical progressions of residues,* Funct. Anal. and Other Math. **2** (2009), no. 2–4, 139–149. MR2506112

[5] ———, *How random are fractional parts of arithmetic progressions?*, Uspekhi Mat. Nauk, **63** (2008), no. 2, 5–20; English transl. in Russian Math. Surveys 63 (2008), no. 2.

[6] ———, *Weak asymptotics of the numbers of solutions of Diophantine equations.* Funktsional. Analiz i Prilozhen. **23** (1999), no. 4, 65–66; English transl. in Functional Anal. Appl. **23** (1999), no. 4. 292–293. MR1746430 (2001k:11190)

[7] R. O. Kuzmin, *On a problem of Gauss*, Dokl. AN SSSR, Ser. A, 1928, 375–380. (See also *Sur un Problème de Gauss.* Atti Congr. Intern. Bologne. 1928. vol. 6. 83–99.)

[8] A. Ya. Khinchine, *Continuous fractions.* Nauka, Moscow, 1978. English transl., Dover, Mineola, NY, 1997. MR514845 (80d:10015)

[9] H. Gylden, *Quelques remarques rélativement à la représentation des nombres irrationelles par des fractions continues*, C. R. Acad. Sci. Paris, **107** (1888), 1584–1587.

[10] M. O. Avdeeva and V. A. Bykovskii, *Solution of the Arnold problem about the Gauss–Kuzmin statistics.* Dalnauka, Vladivostok, 2002. See also Funktsional. Analiz i Prilozhen. **38** (2004), no. 2, 1–11; English transl., Functional Anal. Appl. **38** (2004), no. 2, 79–87. MR2086623 (2005g:11143)

[11] V. I. Arnold, *Arithmetics of binary quadratic forms, symmetry of their continued fractions and geometry of their de-Sitter world*, Bull. Braz. Math. Soc. **34** (2003), no. 1, 1–42. MR1991436 (2004h:11030)

[12] ———, *Statistics of the periods of continued fractions for quadratic irrationals*, Izv. Ross. Akad. Nauk Ser. Mat. **72** (2008) no. 1, 3–38. (Russian) MR2394969 (2009e:11014)

[13] F. Aicardi, *Empirical estimates of the average orders of orbits period lengths in Euler groups*, C. R. Acad. Sci. Paris. Ser. I. **339** (2004), 15–20. MR2075226 (2005f:11001)

[14] V. I. Arnold, *Smooth functions statistics*, Funct. Anal. and Other Math. **1** (2006); see also, Abdus Salam International Centre for Theoretical Physics. ICTP, 2006. IC/2006/012. 9 pp. MR2385493 (2009c:26027)

[15] L. Nicolaescu, *Morse functions statistics*, Funct. Anal. and Other Math. **1** (2006), no. 1, 97–103. MR2381964 (2009g:57055)

[16] V. I. Arnold, *Statistics of the period lengths of the continued fractions for the eigenvalues of the integer matrices of order two*, Funct. Anal. and Other Math. **2** (2007), no. 1, 15–26. MR2466084 (2009k:11010)

[17] H. Tsuchihashi, *Higher-dimensional analogues of periodic continued fractions and cusp singularities*, Tohoku Math. J. **35** (1983), 607–639. MR721966 (86a:14001)

[18] E. I. Korkina, *Two-dimensional continued fractions. The simplest examples*, Trudy Mat. Inst. Steklov. **209** (1995), 143–166. (Russian) MR1422222 (97k:11104)

[19] ———, *La périodicité des fractions continues multidimensionelles*, C. R. Acad. Sci. Paris. **319** (1994), 777–780. MR1300940 (95j:11064)

[20] O. N. Karpenkov, *On examples of two-dimensional continued fractions*, Cahiers du CEREMADE, Université Paris-Dauphine, 2004. No, 0430. 18 pp.

[21] M. L. Kontsevich and Yu. M. Suhov, *Statistics of Klein polyhedra and multidimensional continued fractions*, Pseudoperiodic Topology, V. I. Arnold et al. (eds.) Transl. Amer. Math. Soc., Ser. 2, vol. 197, Providence, RI, 1999, pp. 9–27. MR1733869 (2001h:11101)

[22] V. I. Arnold, *Statistics of Young diagrams of cycles of dynamical systems for finite tori automorphisms*, Moscow Math. J. **6** (2006), no. 1, 43–56. MR2265946 (2007k:05218)

[23] ———, *Experimental observation of mathemaitcal objects*, Summer School "Modern Mathematics", Dubna, 2005. MCCMO, Moscow, 2006. (Russian)

[24] V. L. Goncharov, *On a topic of combinatorial analysis*, Izvestiya AN SSSR, Ser. Matem. **8** (1944), 3–48. (Russian)

[25] I. Percival and F. Vivaldi, *Arithmetical properties of strongly chaotical motions*, Physica D **25** (1987), no. 1, 105–130. MR887460 (88g:58120)

[26] V. I. Arnold, *Statistics of integral convex polygons*, Funtsional. Anal. i ego Prilozhen. **14** (1980), no. 2, 1–3; English transl., Functional Anal. Appl. 14 (1980), no. 2, 79–81. MR575199 (81g:52011)

[27] N. V. Smirnov, *On estimates of divergence of two empirical distribution curves for two independent samples*, Bull. Moskovsk. Universiteta, Matematika, **2** (1939), 3–14. (Russian)

Steklov Mathematical Institute, Russian Academy of Sciences, Moscow, Russia

Translated by O. KHLEBORODOVA