CHAPTER 1

Complex systems

1. Order and Disorder

Mathematics is often called the Science of Order. It is spectacularly successful in creating generalizations: the more-than-2,000-year-old arithmetic and geometry were developed into the monumental fields of calculus, modern algebra, topology, algebraic geometry, and so on. On the other hand, we don't really have a science of disorder. This is why mathematics could say very little about complex systems like the economy or Chess.

What is the fundamental problem with complex systems? The short answer is: the apparent lack of order in the immense space of possibilities. Take, for example, the subject of economics. Economics books, discussing mathematical models (say, von Neumann-type traditional game theory), are always talking about very simple models that quickly settle down into two or three sets of behavior. They are also assuming that the players are infinitely smart and can instantly find the best thing to do in any given situation. But think about what that means in terms of Chess. In the mathematical theory of games there is a general (existential) theorem telling you that any finite two-player game of complete information—such as Chess—has an optimal solution. That is, there is a way of choosing moves that will allow each player (White and Black) to do better than he would with any other choice of moves ("better" more precisely means "not worse")—each player has an *optimal strategy*. In reality, of course, no one has the slightest idea what that optimal strategy is or how to find it.

As I said, Traditional Game Theory (John von Neumann, John Nash, etc.) starts out by saying, "assume rational play", and the ideal rational player the theory talks about could find an optimal strategy instantaneously. Two such ideal rational players, sitting at the opposite sides of an empty chessboard at the start of a play, would just list all the possibilities in their minds and work backward from all possible ways one could force a checkmate. Carrying out such a complete analysis, either player could find an optimal strategy. At that point, there would be no need to actually play the game! Whichever player has the winning strategy—say, White—would immediately claim victory, knowing that he would always win. And the opponent would immediately concede defeat, knowing that he would always lose. If both players have a drawing strategy, then again no one would bother to actually play—the outcome is always a draw, assuming they play rationally. This is the—at first sight absurd—model for game-playing in Traditional Game Theory, due to John von Neumann.

Of course, no human player can carry out a complete analysis of Chess. The estimated total number of moves in Chess is more than 10^{100}, which is more than the number of elementary particles in the observable universe. All that we humble

human players can do is to rely on a few dozen (or hundred, or thousand—depending on the quality of the player) Rules of Thumb, which are hard-learned intuitive guides that tell us what kind of recipes work best in a given situation. This is why the experts call the finite game of Chess an "effectively infinite" game.

Chess is a very complicated game, but the national economy, say, or more narrowly the trade of the USA with Europe or China, is far more complicated than Chess. So here is the fundamental problem of the economy—or Chess, or any other similar complex system—in a nutshell: How can one make a science out of imperfect players exploring their ways into an "effectively infinite" space of possibilities? Is there an escape from the human inability to carry out huge numbers of computations? This is the basic problem of combinatorics and theoretical computer science.

2. Ideal gases and the Equiprobability Postulate

Note that Statistical Mechanics was invented to solve exactly the same kind of fundamental problem. Newton's laws of motion in Classical Mechanics supply a deterministic dynamical model for gases, which are complex collections of molecules in (thermodynamical) equilibrium, bouncing madly off each other and the walls of the container. A mole of gas contains about $6 \cdot 10^{23}$ particles. Due to the huge number of particles, the underlying dynamic is incredibly complicated. In principle the system of differential equations can be solved. In practice it becomes impossible: even using the most powerful computers, the calculations would take more time than the age of the universe. A nondeterministic, statistical approach is the only practical option.

Statistical Mechanics offers a kind of "coarse-grained" representation (the term was invented by the physicist Ehrenfest). What "coarse-grained" means is the following: the system can be described by a single point in the extremely high-dimensional phase space, specifying the positions and the velocities of all molecules, but instead of telling us exactly which point in phase space the system occupies at a given instant, it tells us the *probability* that the point lies in a given region at some instant. To postulate such *a priori* probabilities—usually equiprobability in the phase space—is the basic idea of Statistical Mechanics.

Statistical Mechanics, working with these *a priori* probabilities, can successfully explain why gases have well-defined average properties like density and pressure.

As a simple illustration of Statistical Mechanics, we derive Boltzmann's energy law and the Maxwell–Boltzmann velocity distribution. Consider a monoatomic gas (i.e., all particles are of the same type); the system has N particles. The phase space of the whole system is $6N$-dimensional (3 space coordinates and 3 components of momentum for each particle). The phase space of a single particle is 6-dimensional; we divide it into cells $C_1, C_2, C_3, \ldots$ of equal 6-dimensional volume $|C|$. A point in the phase space of the whole system defines a sequence of positive integers

$$n_1, n_2, n_3, \ldots \quad \text{with} \quad n_1 + n_2 + n_3 + \cdots = N,$$

representing the numbers of particles in cells $C_1, C_2, C_3, \ldots$, respectively; n_1, n_2, n_3, ... are called the "occupation numbers".

Conversely, to any set of positive integers $n_1, n_2, n_3, \ldots$ subject to the condition $n_1 + n_2 + n_3 + \cdots = N$ there corresponds a set $S(n_1, n_2, n_3, \ldots)$ in the

$6N$-dimensional phase space of the whole system which yields these integers as occupation numbers.

The $6N$-dimensional volume of the set $S(n_1, n_2, n_3, \ldots)$ is clearly

$$\frac{N!}{n_1!n_2!n_3!\cdots}|C|^N.$$

Suppose now that the cells C_i are small enough so that the energy of a particle placed at an arbitrary point in the cell C_i is E_i with a negligible error; then we have

$$n_1E_1 + n_2E_2 + n_3E_3 + \cdots = NE,$$

where NE is the total energy of the system.

Boltzmann now finds the vector $(n_1, n_2, n_3, \ldots)$ which maximizes the volume

$$\frac{N!}{n_1!n_2!n_3!\cdots}|C|^N$$

(its logarithm is the "entropy") under the constraints

$$n_1 + n_2 + n_3 + \cdots = N \quad \text{and} \quad n_1E_1 + n_2E_2 + n_3E_3 + \cdots = NE.$$

By introducing the "probabilities" $p_i = n_i/N$ as new variables and applying the weak form of Stirling's formula $K! \approx (K/e)^K = e^{K\log K - K}$, where $\log x$ denotes the natural (base e) logarithm of x, the optimization problem becomes equivalent to the following. We want to minimize the sum $\sum_i p_i \log p_i$ under the constraints $\sum_i p_i = 1$ and $\sum_i p_i E_i = E$. It can be solved by a routine application of the method of Lagrange multipliers. Introduce the auxiliary function with two multipliers, λ and β:

$$F(p_1, p_2, p_3, \ldots; \lambda, \beta) = -\sum_i p_i \log p_i + \lambda\left(\sum_i p_i \ - 1\right) - \beta\left(\sum_i p_i E_i \ - E\right),$$

and compute the partial derivatives $\frac{\partial F}{\partial p_i}$ and set them equal to zero:

$$\frac{\partial F}{\partial p_i} = -\log p_i - 1 + \lambda - \beta E_i = 0,$$

which gives

$$p_i = e^{\lambda - 1} \cdot e^{-\beta E_i}.$$

Since $\sum_i p_i = 1$, we have

$$e^{\lambda-1}\left(\sum_i e^{-\beta E_i}\right) = 1,$$

and so

$$e^{\lambda-1} = \frac{1}{\sum_i e^{-\beta E_i}}.$$

We conclude

$$p_j = \frac{n_j}{N} = \frac{e^{-\beta E_j}}{\sum_i e^{-\beta E_i}}, \tag{1.1}$$

which is the famous Boltzmann energy formula. If we plug in here the kinetic energy $E_i = \frac{1}{2}mv_i^2$, we obtain that the density function of the velocity distribution has the form e^{-cv^2} where $c > 0$ is a fixed constant; this is exactly the familiar normal distribution ("bell curve"). This law—that the velocity distribution of an ideal gas in thermal equilibrium is normal—is called the Maxwell–Boltzmann law.

The denominator in (1.1)

$$Z = \sum_i e^{-\beta E_i} \tag{1.2}$$

is called the *Partition Function* of the system. The Partition Function contains the complete statistical information on the whole system. All macroscopic parameters of the system can be calculated from the Partition Function; for example, the mean energy

$$\overline{E} = \frac{1}{Z} \sum_i E_i e^{-\beta E_i}$$

comes from the derivative of (1.2) with respect to β. The whole existence of the Partition Function is based on the Equiprobability Postulate (meaning that all phase space cells of the same volume are equiprobable).

How can we justify the Equiprobability (or Typicality) Postulate? How does probability enter Classical Mechanics? A well-known attempt to justify the use of *a priori* probabilities—equidistribution—is the so-called Ergodic Hypothesis. It says that the time evolution of a system is a tour that visits all possible states in the phase space in such a way that the time average over a sufficiently long time equals the "ensemble average". Unfortunately, only very special kinds of physical systems ("Sinai billiards" and its variants) are known to have ergodicity, i.e., the time average equals the ensemble average. What is worse, the Kolmogorov–Arnold–Moser theory for small perturbations of integrable (in fact, almost periodic) systems gives us little hope that ergodicity takes place in its original strong form in general physical situations.

The Ergodic Hypothesis, introduced by Boltzmann, led to the developments of a wonderful mathematical theory: Ergodic Theory. One has to admit, however, that Ergodic Theory doesn't seem to help finding a rigorous mathematical foundation for Statistical Mechanics. But this is not such a big problem: the majority of physicists are pragmatists anyway. They are perfectly satisfied with the fact that Statistical Mechanics works: it can correctly predict the outcomes of (most of) the experiments. Agreement with experiment is the best substitute for a rigorous mathematical proof of the Equiprobability Postulate.

Physicists say, "Try this; if it works (with a reasonable level of accuracy), that will justify the postulate". But a mathematician by training is obliged to point out the characteristic fallacy: inductive experience that the postulate works is not a rigorous mathematical proof.

We emphasize that any application of Statistical Mechanics to a concrete problem requires a certain leap of faith. We need to believe that the simplifications are "reasonable"; for example, the ideal gas models simply forget all about intersections between the molecules. In reality, however, the molecules do collide like tiny billiard balls, and actually these collisions give rise to the equilibrium state! (Unfortunately, mathematics cannot handle interacting models.) We should view Statistical Mechanics as an experimental science in the sense that we have to check the theoretical predictions against experiments and hard data.

3. Apparent randomness of primes and the Riemann Hypothesis

We increasingly come to realize that any sufficiently complex phenomenon, even if it is completely deterministic, will often be amenable only to probabilistic

simulation. A particularly rich source of examples is number theory. A good example is the distribution of the prime numbers (I start from 3)

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, \dots.$$

Taking the gap between consecutive primes, we obtain the sequence

$$2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, 2, 4, 6, 6, 2, 6, 4, 2, \dots$$

that shows a highly irregular behavior. In the history of mathematics the set of primes served the first example of what one would call a "random set". Mathematicians have been tantalized by the mystery of primes since the Greeks. The Riemann Hypothesis (a precise statement about the location of the so-called nontrival zeros of the Riemann zeta-function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$—a key point here is that variable s can be extended over the whole complex plane), arguably the most famous open problem in mathematics, is also about the primes, or, at least, it is equivalent to a problem about the "randomness" of the primes. To explain this, first I briefly recall the highlights of prime number theory. The starting point is the unique prime factorization property of the integers $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$, which immediately leads to the product formula ("Euler product")

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{1}{n^s} &= \sum_{n=1}^{\infty} \frac{1}{(p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k})^s} \\ &= \prod_{p:\text{primes}} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots\right) \\ &= \prod_{p:\text{primes}} \frac{1}{1 - p^{-s}}, \end{aligned} \tag{1.3}$$

which holds for every real s with $s > 1$. At $s = 1$ the series $\sum_{n=1}^{\infty} n^{-s}$ becomes divergent and tends to infinity logarithmically:

$$\sum_{n=1}^{x} \frac{1}{n} = \log x + O(1), \tag{1.4}$$

where $\log x$ is the natural logarithm of x. Taking the logarithm of (1.3) and using (1.4), it is not too difficult to obtain the fact that

$$\sum_{p \le x:\text{primes}} \frac{1}{p} = \log\log x + O(1). \tag{1.5}$$

The derivative of $\log\log x$ is $\frac{1}{x \log x}$, so

$$\int_e^x \frac{dt}{t \log t} = \log\log x - \log\log e = \log\log x. \tag{1.6}$$

Comparing (1.5) and (1.6), it is quite natural to come up with the conjecture that the nth prime p_n is asymptotically $n \log n$, or formally, $p_n = (1 + o(1)) n \log n$. So far everything was elementary and rather simple. But the proof of the plausible conjecture $p_n = (1 + o(1)) n \log n$ is anything but simple: it is an alternative form of the deep Prime Number Theorem. The Prime Number Theorem was proved more than 100 years ago; since then the research has focused on the Riemann Hypothesis, which is related to the primes via the remarkable Explicit Formula of Riemann.

What is an "explicit formula" in general? Everybody knows the beautiful formula for the nth Fibonacci number

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right), \tag{1.7}$$

where $F_0 = 0, F_1 = 1, F_2 = 1, F_3 = 2, \ldots$ is the Fibonacci sequence satisfying the recurrence relation $F_{n+2} = F_{n+1} + F_n$ for every $n \geq 0$. The recurrence $F_{n+2} = F_{n+1} + F_n$ motivates the quadratic polynomial $x^2 = x+1$ (called the "characteristic polynomial"), and the key constants $(1+\sqrt{5})/2$, $(1-\sqrt{5})/2$ in (1.7) are the two *zeros* of this polynomial.

The Explicit Formula of Riemann is similar to (1.7) in the sense that it relates the prime-counting function $\pi(x) = \sum_{p \leq x} 1$ (i.e., the number of primes $\leq x$) to the nontrivial *zeros* of the Riemann zeta-function. Instead of the original formula, nowadays it is customary to discuss a simplified version, due to von Mangoldt, where the plain prime-counting function $\pi(x) = \sum_{p \leq x} 1$ is replaced with a weighted version, called "Mangoldt sum". The Mangoldt sum is defined as

$$\psi(x) = \sum_{1 \leq n \leq x} \Lambda(n), \tag{1.8}$$

where $\Lambda(n)$, a logarithmic weight for primes, is defined as follows: $\Lambda(n) = \log p$, if n is a power of p (p always stands for a prime) and $\Lambda(n) = 0$ if n is not a prime-power. (The unusual notation using the Greek letters ζ, π, Λ, and ψ is mandatory in analytic number theory.)

To be precise, (1.8) holds only if x is not a prime-power; if x happens to be a prime-power, then we have to subtract $\Lambda(x)/2$ from the right-hand side of (1.8). The main benefit of the awkward logarithmic weight $\Lambda(n)$ (called the Mangoldt function) is that it has a simple Dirichlet series

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = -\frac{\zeta'(s)}{\zeta(s)}. \tag{1.9}$$

This is how the Riemann zeta-function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ enters the story. We emphasize the key fact that ζ, as a function of the variable s, can be extended over the whole complex plane.

The technical advantage of the Mangoldt function and formula (1.9) is that the logarithmic derivative $\frac{\zeta'(s)}{\zeta(s)}$ only has simple poles, even if $\zeta(s)$ itself happens to have multiple zeros (which, by the way, is not very likely). Thus we have the simple formula

$$-\frac{\zeta'(s)}{\zeta(s)} = \frac{1}{s-1} - \sum_{\rho} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) - \sum_{n=1}^{\infty} \left(\frac{1}{s+2n} - \frac{1}{2n} \right) + O(1), \tag{1.10}$$

where ρ runs through the nontrivial zeta-zeros on the complex plane (meaning the zeros in the vertical strip with real part between 0 and 1). A standard application of Perron's formula (complex integration and residue computation) to the logarithmic derivative $\frac{\zeta'(s)}{\zeta(s)}$ gives the remarkable Explicit Formula in prime number theory:

$$\psi(x) = x - \sum_{\rho} \frac{x^\rho}{\rho} + O(1), \tag{1.11}$$

where again ρ runs through the nontrivial zeta-zeros.

We have very good information about the number of the nontrivial zeta-zeros (say) in the vertical box where the imaginary part has absolute value $\leq T$: the number is

$$\frac{1}{2\pi}T\log T - \frac{1+\log(2\pi)}{2\pi}T + O(\log T). \tag{1.12}$$

A byproduct of (1.12) is that it settles the convergence of (1.11) (if we put the complex conjugate zeta-zeros in pairs).

In sharp contrast to the *number*, we can prove very little about the *location* of the nontrivial zeta-zeros. What we can prove is much, much less than the Riemann Hypothesis, which claims that the nontrivial zeta-zeros are all on the critical line (vertical line with real part 1/2). Applying the Riemann Hypothesis to (1.11), we obtain

$$\psi(x) = x + O(x^{1/2+o(1)}), \tag{1.13}$$

or equivalently (via integration by parts)

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O(x^{1/2+o(1)}). \tag{1.14}$$

The square-root size error term $O(x^{1/2+o(1)})$ nicely fits the well-known "random set" simulation of the primes. By the Prime Number Theorem, the density of the primes at x is $\frac{1}{\log x}$. This motivates the following simulation (due to Cramer; a model that is almost 100 years old): starting from $n = 3$, for every integer $n \geq 3$ we toss a "loaded n-coin" that shows Heads with probability $\frac{1}{\log n}$ and shows Tails with probability $1 - \frac{1}{\log n}$. Keeping n if the outcome of the trial is Heads and rejecting it if the outcome is Tails, we obtain a Random Subset of the natural numbers; we call the elements of this random set "random primes". The expected number of "random primes" is exactly

$$\sum_{n=3}^{x} \frac{1}{\log n} = \int_2^x \frac{dt}{\log t} + O(1), \tag{1.15}$$

and the actual number of "random primes" $\leq x$ fluctuates around the expected number (1.15) with the usual square-root size standard deviation $O(x^{1/2+o(1)})$. In other words, formula (1.14), which is equivalent to the Riemann Hypothesis, is in perfect harmony with the $O(x^{1/2+o(1)})$ size fluctuation of the Random Subset (i.e., a simulation of the primes).

The converse is also true: if the Riemann Hypothesis fails, then the fluctuation in (1.11) is much larger than the standard deviation $O(x^{1/2+o(1)})$. Indeed, if there is a nontrivial zeta-zero $\rho = \beta + i\gamma$ with $\beta \neq 1/2$, then $\rho^* = (1-\beta) + i\gamma$ is another zeta-zero (follows from a symmetry of the Fuctional Equation of the zeta-function), and $\max\{\beta, 1-\beta\} = \alpha > 1/2$. Then in (1.11) the fluctuation around x is at least as large as $x^{\alpha-o(1)}$, and also the fluctuation of $\pi(x)$ around the logarithmic integral is at least as large as $x^{\alpha-o(1)}$, which is asymptotically much larger than the standard deviation $O(x^{1/2+o(1)})$ of the Random Subset (it is not too difficult to make this argument precise). In other words, the failure of the Riemann Hypothesis implies that the "random prime" model is grossly incorrect.

Even if no one has a rigorous mathematical proof, everyone would agree that the Riemann Hypothesis is "true": it cannot be an accident that the first billion zeta-zeros are all on the critical line! Since the Riemann Hypothesis is "true",

we can say that the "random prime" model predicts the fluctuations in the global distribution of primes very accurately.

Also, by using his "random prime" model, Cramer formulated the following daring local conjecture about the maximum gap between consecutive primes:

$$p_{n+1} - p_n = O((\log p_n)^2), \quad \text{or even} \quad p_{n+1} - p_n = O((\log p_n)^{1+o(1)}). \tag{1.16}$$

Numerical evidence supports this conjecture, but, unfortunately, we are very far from proving it. We cannot even prove the much weaker bound $p_{n+1} - p_n = O(p_n^{1/2+o(1)})$.

This was Cramer's probabilistic interpretation of the Riemann Hypothesis (in the early years of the 20th century). About a decade later Denjoy came up with an alternative probabilistic interpretation, which is perhaps even simpler. It is based on the identity (due to Stieltjes)

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, \quad \text{where} \quad \mu(n) = (-1)^k \quad \text{with} \quad n = p_1 \cdots p_k$$

(if n is square-free and has k prime factors); $\mu(n) = 0$ if n is not square-free; $\mu(n)$ is called the Möbius function. The poles of $\frac{1}{\zeta(s)}$ are exactly the zeta-zeros, so it is not surprising that the Riemann Hypothesis is equivalent to the following asymptotic statement:

$$M(x) = \sum_{1 \le n \le x} \mu(n) = O\left(x^{1/2+o(1)}\right) \quad \text{as} \quad x \to \infty. \tag{1.17}$$

This equivalence was proved by Littlewood.

Consider now a very large square-free integer n; then $\mu(n) = \pm 1$. It is "reasonable" to say that $\mu(n)$ is plus or minus one "with equal probability", because, roughly speaking, n will typically have a large number of distinct prime factors, and this number seems to be even or odd "with the same probability". It is perhaps "equally reasonable" to say that the successive evaluations of $\mu(n) = \pm 1$ are "independent". But then the evaluation of the Möbius sum $M(x) = \sum_{1 \le n \le x} \mu(n)$ would be like tossing a fair coin for each square-free $n \le x$ and subtracting the number of Heads from the number of Tails. Since the number of square-free integers $n \le x$ is asymptotically a constant times x (in fact, $6\pi^{-2}x$), the coin-tossing simulation of the Möbius sum $M(x) = \sum_{1 \le n \le x} \mu(n)$ predicts the fluctuation $O(x^{1/2+o(1)})$ ("standard deviation"), which is equivalent to the Riemann Hypothesis (see (1.17)).

Extensive computer studies (mostly due to Odlyzko) show that the nontrivial zeta-zeros also exhibit some kind of a pseudorandomness: the gaps between consecutive zeta-zeros on the critical line are distributed like the gaps between consecutive eigenvalues of random Hermitian (or unitary) matrices.

In other words, the spacing between the zeta-zeros features *rescaled* randomness.

For more about the Riemann zeta-function, see the excellent book of Titchmarsh and Heath-Brown [1986].

4. Zoo of zeta-functions

The Riemann Hypothesis is not an isolated question. Quite the contrary, there is a large variety of zeta-functions, bearing varying degrees of resemblance to the original Riemann zeta-function, each with its own "Riemann Hypothesis". The three most well-known infinite families of zeta-functions are (1) the Dirichlet

L-functions (closely related to quadratic number fields), (2) the Dedekind zeta-functions (based on the unique prime-ideal factorization in arbitrary number fields), and (3) the Hecke L-functions. Each posesses a Functional Equation implying the hidden symmetry, each posesses an associated critical strip, and the corresponding "Riemann Hypothesis" still conjectures that all new nontrivial zeta-zeros lie on the critical line of the original. Again there is overwhelming numerical evidence supporting this so-called "Grand Riemann Hypothesis".

Unfortunately, the resolution of any of these more general Riemann Hypotheses appears just as hard as the original. In spite of the stunning numerical evidence, we are basically clueless about the reasons behind the Grand Riemann Hypothesis.

Finally, note that in Chapters 2 and 3, I will show many more examples of "apparent randomness", supported by numerical evidence but lacking rigorous proof. At the end of Chapter 4, I will summarize the message in a metaphysical conjecture. Starting with Chapter 5, I will give rigorous proofs.

CHAPTER 2

Collecting data: Apparent randomness of digit sequences

1. Normal numbers

Consider the usual decimal expansion of real numbers, and start with the following three statements: (1) "in the decimal expansion of $\sqrt{2}$, 7 has density 1/10", (2) "in the decimal expansion of $e = 2.71828\ldots$, 7 has density 1/10", and (3) "in the decimal expansion of $\pi = 3.14159\ldots$, 7 has density 1/10". Which one is true? Are they all true? Well, they are all famous unsolved problems. Despite the empirical fact that, in all three cases, among the first million decimal digits there is a solid computational evidence for equidistribution of the digits, we cannot rule out the possibility that, say, the density in question does not exist. Computer experimentation is not a rigorous mathematical proof.

From the digit 7 we can switch to any other digit and also to any block of consecutive digits. In general, a real number is said to be *normal* in a particular base (say, the standard decimal case $b = 10$) if every block of digits of any length occurs with equal density (depending only on the length of the block: the density is b^{-l} if the length is l and the base is b). Ever since early computer experiments in the 1940s, enormous empirical evidence has been accumulating for the normality of special numbers like $\sqrt{n}$ (where n is not a square), $e, \pi, \log 2, \log 3$, cube roots, fourth roots, etc. In spite of this numerical evidence, no "special" number has ever been rigorously proved to be normal. More precisely, no number expressed just in terms of standard mathematical functions has been proved to be normal.

We know that almost all real numbers are normal—this is a classical theorem of Borel from 1909. But just because *almost all* real numbers are normal, and, what is more, almost all reals are normal in *all* bases, it is absolutely not clear why the individual numbers like $\sqrt{2}$ and e and π are all normal without exception.

The rational numbers are of course trivial exceptions (periodicity contradicts normality). On the other hand, it is not easy to construct an explicit normal number. Borel himself couldn't do it; the first example was constructed only in 1933. This example is shown below; it is normal in base 10:

$$.12345678910111213141516171819202 1\ldots 99100101102\ldots.$$

The simple rule is that the digits are those of all natural numbers in succession ("concatenation sequence"). Unfortunately this number, called the Champernowne number, is not a "special" number like $\sqrt{2}$ or e or π. Is the Champernowne number normal in base 2? How about base 3?

Even now, 100 years after Borel's theorem, we still don't know any explicit example of a number that is normal in all bases (though $\sqrt{2}$, π, and e are all very likely candidates, supported by overwhelming numerical evidence).

2. Continued fraction

Unlike ordinary decimal digits, the individual terms in a continued fraction (often called *partial quotients*) can be of any size. For example, we know from Euler that

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cdots}}}}},$$

or by using the space-saving notation,

$$e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, \ldots, 1, 2n, 1, \ldots]. \tag{2.1}$$

Notice that (2.1) has a simple linear pattern.

The size of the partial quotients plays a key role in diophantine approximation. As an illustration consider the approximation $\pi \approx \frac{355}{113}$ (known to a Chinese mathematician more than 1,500 years ago!), which is accurate to the first six decimal places. The reason behind this good approximation is that in the continued fraction for π,

$$\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, \ldots],$$

an unusually large term—namely 292—shows up at a very early stage. Note that $\frac{355}{113} = [3; 7, 15, 1]$ is the 4th convergent of π.

To represent a real number $0 < x < 1$ as a continued fraction, we take the reciprocal $1/x$ and write it as the sum of the lower integral part $a_1 = \lfloor 1/x \rfloor$ and the fractional part $0 < \{1/x\} < 1$. Now we do to $x_1 = \{1/x\}$ what was done to x and keep repeating the process:

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cfrac{1}{a_5 + \cdots}}}}}. \tag{2.2}$$

The process is infinite, unless x is rational. This motivates the function $T: x \to \{1/x\}$, which maps the interval (0,1) onto itself. Note that T is *not* one-to-one: the inverse image of an interval (a, b), where $0 < a < b < 1$, is the infinite union of disjoint intervals

$$\left(\frac{1}{1+b}, \frac{1}{1+a}\right), \left(\frac{1}{2+b}, \frac{1}{2+a}\right), \left(\frac{1}{3+b}, \frac{1}{3+a}\right), \cdots; \tag{2.3}$$

each one of these intervals is mapped to the whole of (a, b) by T.

If we define the measure of an interval (a, b) to be

$$m(a, b) = \frac{1}{\log 2} \log \frac{1+b}{1+a} = \frac{1}{\log 2} \int_a^b \frac{dx}{1+x}, \tag{2.4}$$

then one can easily check that this m-measure of the interval (a, b) equals the sum of the m-measures of the intervals in (2.3). We can extend (2.4) to any measurable

set $A \subset (0,1)$ by the integral

$$m(A) = \frac{1}{\log 2} \int_A \frac{dx}{1+x}. \tag{2.5}$$

Measure defined in (2.4)–(2.5) was already known to Gauss (who carried out an extensive numerical experimentation on continued fractions). The key property of measure (2.4)–(2.5) is that it is preserved by the transformation T. By (2.2) the first partial quotient a_1 of a real $x \in (0,1)$ equals an integer $k \geq 1$ if and only if x falls into the interval $(\frac{1}{k+1}, \frac{1}{k})$, which has m-measure

$$\frac{1}{\log 2} \int_{1/k+1}^{1/k} \frac{dx}{1+x} = \frac{1}{\log 2} \left(\log(1+\frac{1}{k}) - \log(1+\frac{1}{k+1}) \right) = \frac{\log \frac{(k+1)^2}{k(k+2)}}{\log 2}. \tag{2.6}$$

A well-known theorem of Kusmin states that, for almost all $x \in (0,1)$, the density with which an arbitrary integer $k \geq 1$ appears in the sequence $a_1, a_2, a_3, \ldots$ of partial quotients in (2.2) is exactly (2.6). For example, for almost all $x \in (0,1)$, the density of the digit 1 is exactly

$$\frac{\log(4/3)}{\log 2} = .415\ldots \approx 41.5\%. \tag{2.7}$$

It was realized later that Kusmin's theorem is a special case of the very general Ergodic Theorem of Birkhoff. (Note, however, that Birkhoff's general theorem doesn't give any error term, but in Kusmin's theorem we can prove the best possible error term.)

Kusmin's theorem clearly fails for $x = e$: by (2.1) the frequency of digit 1 is $2/3$, which differs from the 41.5% in (2.7). Note that a few "relatives" of e have a pattern similar to (2.1). For example,

$$\sqrt{e} = [1; 1, 1, 1, 5, 1, 1, 9, 1, 1, 13, 1, \ldots, 1, 4n+1, 1, \ldots],$$
$$e^2 = [7; 2, 1, 1, 3, 18, 5, 1, 1, 6, 30, \ldots, 3n-1, 1, 1, 3n, 12n+6, \ldots];$$

they all violate Kusmin's theorem.

The most famous violators of Kusmin's theorem are the quadratic irrationals, such as $\sqrt{2}$ and $\sqrt{3}$. The general form is $(a+\sqrt{b})/c$ where a, $b \geq 2$, $c \geq 1$ are integers, and b is not a square. By a well-known result in number theory, all quadratic irrationals have periodic continued fractions, and the converse is also true: periodic continued fractions all arise from quadratic irrationals. Here are a few examples:

$$\sqrt{2} = [1; 2, 2, 2, 2, 2, \ldots] = [1; \overline{2}],$$
$$\sqrt{3} = [1; 1, 2, 1, 2, 1, 2, 1, 2, \ldots] = [1; \overline{1, 2}],$$
$$\sqrt{67} = [8; \overline{5, 2, 1, 1, 7, 1, 1, 2, 5, 16}],$$
$$\sqrt{94} = [9; \overline{1, 2, 3, 1, 1, 5, 1, 8, 1, 5, 1, 1, 3, 2, 1, 18}];$$

incidentally $\sqrt{94}$ has the longest period among the first hundred integers. By contrast, higher roots (cube roots, fourth roots, etc.) never appear to show any simple patterns like what e or $\sqrt{e}$ or e^2 does. Unlike "regularity", they all seem to show "randomness" with Kusmin's rescaling (see (2.6)).

For example, among the first million partial quotients in the continued fraction for the cube root of 2, the digit 1 appears 414,983 times, which is remarkably close to the 41.5% in (2.7), i.e., Kusmin's limit (2.6) with $k = 1$.

The same remarkable fact holds for the special number π: among the first million partial quotients, the digit 1 appears 414,526 times—very close to 41.5%.

These are striking numerical facts, but, unfortunately, we cannot prove any theorem—not even the most plausible conjecture. For example, we don't know for sure whether the sequence $a_1, a_2, a_3, \ldots$ of partial quotients for the cube root of 2 is bounded or not. What is worse, we don't know a single algebraic number of degree ≥ 3 for which the sequence $a_1, a_2, a_3, \ldots$ of partial quotients is unbounded. We don't know this in spite of the well-known conjecture claiming that $a_1, a_2, a_3, \ldots$ is unbounded for *every* single real algebraic number of degree ≥ 3.

Computer experimentation seems to indicate that the continued fraction for a special number is either obviously simple (like for the quadratic irrationals or for e and its relatives) or it shows randomness with Kusmin's rescaling.

This is exactly the same message that appears to be the case for the ordinary decimal expansion (normal numbers). The only technical difference is in *rescaling*: in continued fractions the ordinary uniform Lebesgue measure in the interval (0,1) has to be replaced by the nonuniform measure (2.4)–(2.5).

3. Equidistribution and continued fraction

The starting point of the theory of Uniform Distribution was the discovery that the sequence $n\alpha$, $n = 1, 2, 3, \ldots$ modulo one, is uniformly distributed for every irrational α (Bohl, Sierpinski, Weyl; early 1900s). Of course the rational numbers are all violators: the modulo one sequence is periodic. It is natural, therefore, to measure how well one can approximate an irrational α with rationals of small denominator. The size of the partial quotients a_n in the continued fraction for α tells us the whole story. Indeed, if

$$\alpha = [a_0; a_1, a_2, a_3, \ldots] \quad \text{and} \quad \frac{p_n}{q_n} = [a_0; a_1, a_2 \ldots, a_n],$$

then $p_n = a_n p_{n-1} + p_{n-2}$, $q_n = a_n q_{n-1} + q_{n-2}$,

$$\alpha = \frac{p_n}{q_n} + \frac{(-1)^n}{q_n(q_{n+1} + \theta q_n)} \quad \text{for some} \;\; 0 < \theta < 1,$$

and so

$$\left| \alpha - \frac{p_n}{q_n} \right| \approx \frac{1}{q_n q_{n+1}} \approx \frac{1}{q_n^2 \cdot a_{n+1}}. \tag{2.8}$$

We can roughly say that the smaller the partial quotients are (i.e., α is badly approximable by rationals) the more uniform the sequence $n\alpha$ (mod 1) is. It is not too hard to turn this intuition into a precise quantitative formula expressing the discrepancy in terms of the partial quotients a_n. This was done by Ostrowski, Hardy, and Littlewood around 1920.

Here I insert a remark about real algebraic numbers (say, the cube root of 2). Computer experimentation seems to support the fact that the partial quotients satisfy the upper bound

$$a_n = O\left(n^{O(1)}\right), \quad \text{in fact} \;\; a_n = O\left(n^{1+o(1)}\right); \tag{2.9}$$

at least this is the case for every single real algebraic number ever tried. Assuming a Kusmin-type rescaled "randomness", conjecture (2.9) is perfectly plausible. It is plausible the same way as Cramer's conjecture about the maximum gap between

consecutive primes is plausible; see (1.16). Unfortunately, what we can prove to date is *far* weaker (see (2.19) below).

The theory of "irrational rotation" $n\alpha \pmod 1$ is the most complete chapter of Uniform Distribution (I will return to it in Chapter 5). By contrast, we know very little about exponential sequences modulo one. It has been known since the 1930s that the sequence α^n, $n = 1, 2, 3, \ldots$, is uniformly distributed modulo one for almost all real numbers $\alpha > 1$ (an analog of Borel's theorem, proved by Koksma). However, no specific value of $\alpha > 1$ for which this is true has ever been explicitly found!

The most famous special case is $\alpha = 3/2$. Computer experimentations suggest that $(3/2)^n$ is indeed uniformly distributed modulo one, but despite a fair amount of mathematical work since the 1940s, there has been no real progress towards proving this.

A typical exception is $\alpha = 1 + \sqrt{2}$ (for the simple reason that, with $\beta = 1 - \sqrt{2}$, $\alpha^n + \beta^n$ is an integer and $\beta^n \to 0$); the class of similarly defined numbers is called Pisot numbers. Is it true that every exception is a Pisot number?

4. More on continued fraction and diophantine approximation

Let's return to (2.8). It is well known that the convergents p_n/q_n of a real α (for simplicity assume that α is irrational) give the "best" rational approximations of α. Formula (2.8) says that the larger a_{n+1} the better the approximation.

The diophantine approximation behavior of almost all α (in the sense of the usual Lebesgue measure) is relatively simple, and it is completely described by the following

KHINTCHIN'S THEOREM. *If the function $f(q)$ increases with q and the series*

$$\sum_{q=1}^{\infty} \frac{1}{qf(q)}$$

is divergent, then for almost all α the diophantine inequality

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2 f(q)} \tag{2.10}$$

is solvable for infinitely many integers q.

Note that the converse is also true (it has a much simpler proof). The easy converse states that, if

$$\sum_{q=1}^{\infty} \frac{1}{qf(q)}$$

is convergent, then for almost all α the diophantine inequality (2.10) has only a finite number of solutions p/q.

For example, if $f(q) = (\log q)^c$ with some fixed constant $c > 0$, then the series

$$\sum_{q=1}^{\infty} \frac{1}{qf(q)}$$

is divergent for $0 < c \le 1$, and so, for almost all α, (2.10) has infinitely many solutions. On the other hand, the series is convergent for $c > 1$, and so, for almost all α, (2.10) has only a finite number of solutions.

Now let's leave the easy case of almost all α and assume that α is an arbitrary fixed real *algebraic* number of degree $k \geq 2$. Then we have the basically trivial inequality

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c(\alpha)}{q^k} \tag{2.11}$$

that holds for any rational number p/q, where $c(\alpha) > 0$ is an absolute constant depending only on α. By using this simple observation, around 1844 Liouville was able to show that some extremely rapidly convergent series, such as

$$\sum_{n=1}^{\infty} 2^{-n!} \quad \text{or} \quad \sum_{n=1}^{\infty} 3^{-n^n},$$

represent numbers that are *not* algebraic. These numbers, called Liouville numbers, were the first examples of provably transcendental numbers, and a new theory was born.

Sixty years later Liouville's basic idea was greatly extended by A. Thue. In 1909 Thue was able to prove the first general theorem in the theory of diophantine equations (which initiated a most important line of research in 20th-century number theory). Thue was studying the class of homogeneous polynomials of degree k in two variables,

$$G(x, y) = a_0 x^k + a_1 x^{k-1} y + a_2 x^{k-2} y^2 + \cdots + a_k y^k, \tag{2.12}$$

and asked the question: what is the number of integral solutions $(x, y) \in \mathbb{Z}^2$ of the diophantine equation

$$G(x, y) = b? \tag{2.13}$$

Of course, we assume that a_0, $a_1, \ldots$, a_k, b are all integers, and we also assume, for simplicity, that the polynomial

$$g(x) = G(x, 1) = a_0 x^k + a_1 x^{k-1} + a_2 x^{k-2} + \cdots + a_k \tag{2.14}$$

is irreducible over the rationals. For degree $k = 2$, (2.12) includes the Pell equation, so the number of integral solutions of (2.13) can certainly be infinite. But, if the degree $k \geq 3$, then by Thue's famous theorem, the number of integral solutions of (2.13) is always finite!

I recall Thue's argument in a nutshell. Assume that $G(x, y) = b$ has infinitely many integral solutions. Then there is a real zero α of the polynomial g (see (2.14)) and there are infinitely many integral solutions $x = p_n$ and $y = q_n$ $(n = 1, 2, 3, \ldots)$ of $G(x, y) = b$ such that

$$\left|\alpha - \frac{p_n}{q_n}\right| \leq \frac{c}{q^k} \tag{2.15}$$

holds for some fixed constant c depending only on the coefficients a_0, $a_1, \ldots$, a_k and on b.

Notice that α is a real algebraic number of degree $k \geq 3$. The difficult part in Thue's argument is the proof of the following technical improvement on Liouville's (trivial) bound (2.11).

Thue's Diophantine Approximation Theorem. *If α is a real algebraic number of degree $k \geq 3$, then for any $\varepsilon > 0$,*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c(\alpha, \varepsilon)}{q^{1+\varepsilon+(k/2)}} \tag{2.16}$$

holds for any rational number p/q, where $c(\alpha,\varepsilon) > 0$ is an absolute constant depending only on α and ε.

The point here is that (2.15) and (2.16) contradict each other if $1+\varepsilon+(k/2) < k$, that is, if $\varepsilon < (k-2)/2$. Since $k \geq 3$ and $\varepsilon > 0$ in (2.16) can be arbitrarily small, we can easily enforce this contradiction, and it proves that the number of integral solutions has to be finite.

The proof of (2.16) is very obscure: it is based on the construction of an auxiliary polynomial in two variables of very high degree possessing zeros to a high order. There is no *a priori* explanation for the introduction of this auxiliary polynomial.

Even if we don't understand why it works, Thue's auxiliary polynomial technique does work remarkably well and admits far-reaching generalizations and improvements. The exponent $1+\varepsilon+(k/2)$ in (2.16) seems to be rather "accidental", and it is natural to expect large improvements from a switch to several variables. And indeed, along these lines, in 1955 K.F. Roth managed to replace the "accidental" exponent $1+\varepsilon+(k/2)$ with the best possible exponent $2+\varepsilon$ in (2.16) (notice that $2+\varepsilon$ is basically independent of the degree k; there is only a minor dependence in $\varepsilon = o(1)$).

Roth's Diophantine Approximation Theorem. *If α is a real algebraic number of degree $k \geq 3$, then for any $\varepsilon > 0$,*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c(\alpha,\varepsilon)}{q^{2+\varepsilon}} \tag{2.17}$$

holds for any rational number p/q, where $c(\alpha,\varepsilon) > 0$ is an absolute constant depending only on α and ε.

Note that Roth's theorem was extended to the wider question concerning the *simultaneous* approximation of (linearly independent) algebraic numbers by rationals. This highly nontrivial task was accomplished by W. M. Schmidt in 1970. Schmidt's theorem has striking applications in some classes of diophantine equations of several variables ("norm form equations").

It is worthwhile to compare the extremely difficult Roth's theorem to the very simple metrical case of "almost all α": for any $\varepsilon > 0$, almost all α have the property that

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c(\alpha,\varepsilon)}{q^2(\log q)^{1+\varepsilon}} \tag{2.18}$$

holds for every rational number p/q, where $c(\alpha,\varepsilon) > 0$ is an absolute constant depending only on α and ε. In 1965 S. Lang made the bold conjecture that in Roth's theorem the extra factor $q^{\varepsilon} = q^{o(1)}$ in the denominator can be replaced by $(\log q)^{1+\varepsilon}$; that is, the metrical result (2.18) also holds for every single real algebraic number.

Note that now, forty years later, Lang's conjecture remains just as hopeless. The only tool that we have is the Thue method, and its parameters were apparently optimized by Roth. Roth's complicated proof gives very slow convergence $\varepsilon = o(1) \to 0$ in terms of q as $q \to \infty$; the convergence is certainly slower than $1/\log\log\log q$. In view of this slow convergence, there remains a basically exponential gap between Roth's theorem and Lang's conjecture.

This exponential gap is expressed in a more transparent way in terms of the continued fraction "digits"—called partial quotients—of an algebraic number α.

From Roth's theorem we cannot deduce more than the weak upper bound

$$a_n = O(e^{o(n)}) \quad \text{for the } n\text{th partial quotient of } \alpha. \tag{2.19}$$

On the other hand, it is widely believed that

$$a_n = O\left(n^{1+o(1)}\right)$$

holds for every single real algebraic number; see (2.9).

(I didn't mention a fundamental handicap of the Thue method: ineffectiveness. This handicap prevents us from determining the complete set of solutions of many diophantine equations for which we otherwise know that the number of solutions is finite. For Thue's equation, however, A. Baker did find an effective solution.)

The reader who is interested in reading more about discrete systems based on numbers can find a lot of interesting material in Wolfram's book [2002] (see Chapter 4) and in Klee–Wagon [1991] (see Chapter 3). On uniform distribution and diophantine approximation, see e.g. Drmote–Tichy [1997] and Schmidt [1980]. Also, I strongly recommend to the reader Terence Tao's ICM/2006 lecture on structure and randomness; see T. Tao [2006].