

# Preface

We could choose *randomness of  $\sqrt{2}$*  as an alternative subtitle of the book. Indeed, the book connects two seemingly unrelated concepts, namely, (1)  $\sqrt{2}$ : symbolizing the class of quadratic irrationals, including the theory of the quadratic number fields in general and (2) randomness. These two concepts, representing algebra (the science of order and structure) and probability theory (the science of disorder), are the endpoints of a long chain of relations/implications. The periodicity of the continued fraction of  $\sqrt{2}$  (or any other quadratic irrational) means self-similarity. Self-similarity leads to independence (e.g., via Markov chains; here we refer to the well known probabilistic concept), and independence ensures (nearly) perfect randomness. In particular, we prove some unexpected probabilistic results:

quadratic irrational  $\implies$  periodic continued fraction  $\implies$   
 $\implies$  self-similarity  $\implies$  independence (or independence via Markov chains)  $\implies$   
 $\implies$  randomness : central limit theorem and the law of the iterated logarithm

This diagram may summarize the book in a nutshell.

The reason why we decided not to choose *randomness of  $\sqrt{2}$*  to be the subtitle is that it would perhaps mislead the reader. The reader would probably expect us to prove the apparent randomness of the digit distribution in the usual decimal expansion

$$\sqrt{2} = 1.414213562373095048801688724209698078569671875376948 \dots$$

Unfortunately, we cannot make any progress with this famous old problem; it remains open and hopeless (to read more about this and other related famous open problems the reader may jump ahead right now to Sect. 2.5: A Giant Leap in number theory). What we study instead is the “irrational rotation” by any

quadratic irrational, say, by  $\sqrt{2}$ . We study the global and local behavior of the irrational rotation from a probabilistic viewpoint—this explains the title of the book *probabilistic diophantine approximation*.

Consider the linear sequence  $n\alpha$ ,  $n = 1, 2, 3, \dots$ : it is perfectly regular, it is an infinite arithmetic progression. Even if we take it modulo one, and  $\alpha$  is an arbitrary (but fixed) irrational, the sequence  $n\alpha \pmod{1}$ —called irrational rotation—still features a lot of regularities. For example, (1) we have infinitely many *Bounded Error Intervals*, (2) we have infinitely many *Bounded Error Initial Segments*, (3) every initial segment has at most three different “gaps,” and (4) there is an extremely strong restriction on the induced permutations—these are all strong “anti-randomness” type regularity properties of the irrational rotation  $n\alpha \pmod{1}$ ,  $n = 1, 2, 3, \dots$  (properties (1)–(4) will be explained in depths in Sect. 1.1). These regularities show that the irrational rotation is highly non-random in many respects. This is why the irrational rotation (with an underlying nested structure) is also called a quasi-periodic sequence.

Also we know from number theory that the key to understand the irrational rotation  $n\alpha \pmod{1}$ ,  $n = 1, 2, 3, \dots$ , is to know the continued fraction for  $\alpha$ . The quadratic irrationals have the most regular continued fraction: the class of quadratic irrationals is characterized by the property of (ultimately) periodic continued fraction, for example,

$$\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2 + \dots}} = [1; 2, 2, 2, \dots] = [1; \overline{2}].$$

Despite these regularities of the irrational rotation, our first main result exhibits “full-blown randomness.” For example, how much time does the irrational rotation  $n\alpha \pmod{1}$ ,  $n = 1, 2, 3, \dots$ , spend in the first half  $[0, 1/2)$  of the unit interval  $[0, 1)$ ? Well, we prove a central limit theorem for every quadratic irrational  $\alpha$  (e.g.,  $\alpha = \sqrt{2}$ ). More precisely, let  $\alpha$  be an arbitrary real root of a quadratic equation with integer coefficients, say,  $\alpha = \sqrt{2}$ . Given any rational number  $0 < x < 1$  (say,  $x = 1/2$ ) and any positive integer  $n$ , we count the number of elements of the sequence  $\alpha, 2\alpha, 3\alpha, \dots, n\alpha$  modulo 1 that fall into the subinterval  $[0, x]$ . We prove that this counting number satisfies a central limit theorem in the following sense. First, we subtract the “expected number”  $nx$  from the counting number and study the typical fluctuation of this difference as  $n$  runs in a long interval  $1 \leq n \leq N$ . Depending on  $\alpha$  and  $x$ , we may need an extra additive correction of constant times logarithm of  $N$ ; furthermore, what we always need is a multiplicative correction: division by (another) constant times square root of logarithm of  $N$ . If  $N$  is large, the distribution of this renormalized counting number, as  $n$  runs in  $1 \leq n \leq N$ , is very close to the standard normal distribution (bell-shaped curve), and the corresponding error term tends to zero as  $N$  tends to infinity. This is one of the main results of the book (see Theorem 1.1). The proof is rather complicated and long; it has many interesting detours and by-products. For example, the exact determination of the

key constant factors (in the additive and multiplicative norming), which depend on  $\alpha$  and  $x$ , requires surprisingly deep algebraic tools such as Dedekind sums, the class number of quadratic fields, and generalized class number formulas.

Perhaps the reader is wondering: why are the quadratic irrationals (like  $\sqrt{2}$ ) special and worth spending hundreds of pages on. The answer is that the quadratic irrationals play a central role in diophantine approximation for several reasons. They are the “most anti-rational real numbers” (officially called *badly approximable numbers*), and at the same time they represent the most uniformly distributed irrational rotations. A third reason is the Pell’s equation  $x^2 - dy^2 = \pm 1$  ( $d \geq 2$  is square free), which is of course closely related to  $\sqrt{d}$ . Also, and this is the message of our book, the best way to understand the local and global randomness of the irrational rotation is to focus on the class of quadratic irrationals. This class gives the most elegant and striking results with the simplest proofs. Some of these results extend to *almost every real number*, some of them do not extend. We will elaborate on each one of these issues later.

The quadratic irrational rotation demonstrates the coexistence of order and randomness; a novelty here is the *much* smaller norming factor  $\sqrt{\log n}$  (instead of the usual  $\sqrt{n}$ ). The  $\log n$  comes from the fact that the underlying problem is about “generalized digit sums” with the surprising twist that the base of the number system is an irrational number (namely, the fundamental unit, e.g., it is  $1 + \sqrt{2}$  for  $\alpha = \sqrt{2}$ ). Also  $\log n$  represents the minimum; it corresponds to the most uniformly distributed irrational rotations.

Our second main subject is motivated by the classical Pell’s equation. Finding the integral solutions of (say)  $x^2 - 2y^2 = \pm 1$  means counting lattice points in a long and narrow tilted hyperbolic region that we call a “hyperbolic needle.” Of course, we basically know everything about Pell’s equation (this is why Pell’s equation is included in every undergraduate number theory course), but what happens if we translate the “hyperbolic needle”? What is the asymptotic number of lattice points inside (note that the area is infinite)? Well, for a typical translated copy of the “hyperbolic needle”—which corresponds to an “inhomogeneous Pell inequality”—we prove a “law of the iterated logarithm,” which describes the asymptotic number of integral solutions in a strikingly precise way. In other words, the classical Circle Problem of Gauss is wide open, but here we can solve an analogous Hyperbola Problem. This result is a good illustration of the full power of the probabilistic viewpoint in number theory. In general, consider the inhomogeneous diophantine inequality

$$\|n\alpha - \beta\| < \frac{c}{n}, \quad (0.1)$$

where  $\alpha$  is an arbitrary irrational,  $\beta, c > 0$  are arbitrary real numbers, and  $n$  is the variable. An old result of Kronecker states that inequality (0.1) has infinitely many integral solutions  $n$  if  $c = 3$ ; this is how Kronecker proved that the irrational

rotation  $n\alpha \pmod{1}$  is dense in the unit interval. What can we say about the number of solutions  $n$  of inequality (0.1)? Consider the special case  $\alpha = \sqrt{2}$  of (0.1):

$$\|n\sqrt{2} - \beta\| < \frac{c}{n}, \quad (0.2)$$

and let  $\mathcal{F}(\sqrt{2}; \beta; c; N)$  denote the number of integral solutions  $n$  of inequality (0.2) satisfying  $1 \leq n \leq N$ ; this counting function is about the local behavior of the irrational rotation  $n\sqrt{2} \pmod{1}$ . We can describe the true order of  $\mathcal{F}(\sqrt{2}; \beta; c; N)$ , as  $N \rightarrow \infty$ , in an extremely precise way for almost every  $\beta$ . We prove that the number of solutions  $\mathcal{F}(\sqrt{2}; \beta; c; e^n)$  of (0.2) oscillates between the sharp bounds ( $\varepsilon > 0$ )

$$2cn - \sigma\sqrt{n}\sqrt{(2+\varepsilon)\log\log n} < \mathcal{F}(\sqrt{2}; \beta; c; e^n) < 2cn + \sigma\sqrt{n}\sqrt{(2+\varepsilon)\log\log n} \quad (0.3)$$

as  $n \rightarrow \infty$  for *almost every*  $\beta$ ; see Theorem 5.6 in Part 1.3 of the book. Note that  $\sigma = \sigma(\sqrt{2}, c) > 0$  is a positive constant, and (0.3) fails with  $2 - \varepsilon$  instead of  $2 + \varepsilon$ . (The reason why in (0.3) we switched from  $N$  to the exponentially sparse sequence  $e^n$  is that the counting function  $\mathcal{F}(\sqrt{2}; \beta; c; N)$  is slowly changing in the sense that, as  $N$  runs in  $e^n < N < e^{n+1}$ ,  $\mathcal{F}(\sqrt{2}; \beta; c; N)$  makes only an additive constant change.)

Observe that inequality (0.2) is (basically) equivalent to the inhomogeneous Pell inequality

$$-c' \leq (x + \beta)^2 - 2y^2 \leq c', \quad (0.4)$$

where  $c' = 2\sqrt{2}c$ . Notice that equation (0.4) determines a long and narrow tilted hyperbola region (“hyperbolic needle”). The message of (0.3) is, roughly speaking, that for almost all translations, the number of lattice points in long and narrow hyperbola segments of any fixed quadratic irrational slope equals the area plus an error term which is *never* much larger than the square root of the area.

Notice that (0.3) is a perfect analog of Khinchin’s law of the iterated logarithm in probability theory (describing the maximum fluctuations of the digit sums of a typical real number  $\beta$ ; the factor  $\log\log n$  in (0.3) explains the name “iterated logarithm”).

We also have an analogous central limit theorem: the renormalized counting function

$$\frac{\mathcal{F}(\sqrt{2}; \beta; c; e^n) - 2cn}{\sigma\sqrt{n}}, \quad 0 \leq \beta < 1,$$

has a standard normal limit distribution with error term  $O(n^{-1/4}(\log n)^3)$  as  $n \rightarrow \infty$  [ $\sigma = \sigma(\sqrt{2}, c) > 0$  is the same positive constant as in (0.3)].

Formally,

$$\max_{\lambda} \left| \text{measure} \left\{ \beta \in [0, 1) : \mathcal{F}(\sqrt{2}; \beta; c; e^n) - 2cn \geq \lambda \sigma \sqrt{n} \right\} \right. \\ \left. - \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-u^2/2} du \right| = O_{\gamma} \left( n^{-1/4} (\log n)^3 \right), \quad (0.5)$$

where the maximum is taken over all  $-\infty < \lambda < \infty$  (and of course measure means the one-dimensional Lebesgue measure).

The proofs of the innocent-looking results (0.3) and (0.5) are quite difficult (in spite of the fact that most of the arguments are “elementary”). Note that here “independence” comes from a good approximation by modified Rademacher functions.

The book is basically “lattice point counting” in disguise. This explains the subtitle *randomness in lattice point counting*. The main results are proved by the same scheme: we represent a natural lattice point counting function in the form

$$X_1 + X_2 + X_3 + \dots + \text{negligible},$$

where  $X_1, X_2, X_3, \dots$  are independent random variables. This way we can directly apply some classical results of probability theory (such as the central limit theorem and the law of the iterated logarithm). We have the following questions: (a) how to construct the independent random variables  $X_1, X_2, X_3, \dots$ , (b) how to compute the expectation, and finally (c) how to compute the variance. These are surprisingly difficult questions.

Of course (0.3) and (0.5) extend to all quadratic irrationals. They also extend to some other special numbers for which we know the continued expansion (e.g.,  $e, e^2, \sqrt{e}$ ).

Some of the main results about quadratic irrationals (e.g., Theorems 1.1 and 1.2) do not extend to *almost every*  $\alpha$ . The reason is that the continued fraction digits (officially called partial quotients) of a *typical* real number  $\alpha$  exhibit a very irregular behavior (see Sect. 6.10).

Some other results, including (0.3) and (0.5), do have an analog for *almost every*  $\alpha$ . There is, however, a difference: the norming factor  $\sqrt{n}$  is replaced by  $\sqrt{n \log n}$ , and also the error term is much weaker (see Sect. 6.10).

The kind of “randomness” we prove in the book requires some knowledge about the continued fraction expansion of the real number  $\alpha$ . This is why the best way to demonstrate this “randomness” is to study the class of quadratic irrationals. Unfortunately, we know very little about the continued fraction of algebraic numbers of degree  $\geq 3$ . This explains why we cannot prove anything about (say) the “randomness of  $\sqrt[3]{2}$ ”; this is why we can prove strong results about the “randomness of  $e$ ,” and can prove nothing about the “randomness of  $\pi$ .”

Besides “randomness,” the other main subject of the book is “Area Principle versus superirregularity” (see Part 1.3, starting with Sect. 5.1).

The traditional meaning of *probabilistic diophantine approximation* is that it is a collection of results best illustrated by the following classical 0–1 law of Khinchin. If  $\psi(n) > 0$  is a nonincreasing sequence, then the diophantine inequality  $n\|n\alpha\| < \psi(n)$  has infinitely many integral solutions  $n$  for almost every  $\alpha$  if  $\sum_{n=1}^{\infty} \psi(n) = \infty$ ; on the other hand, if  $\sum_{n=1}^{\infty} \psi(n) < \infty$  then  $n\|n\alpha\| < \psi(n)$  has only finitely many integral solutions  $n$  for almost every  $\alpha$ .

The subtitle of our book (randomness in lattice point counting) emphasizes the fact that what we do here is very different. We develop a new direction of research on the borderline of probability theory and number theory (including algebraic number theory). We switch the focus from almost every  $\alpha$  to special numbers (like quadratic irrationals and  $e$ ), and switch from 0–1 laws to more sophisticated probabilistic results such as the central limit theorem and the law of the iterated logarithm.

One of the challenges we faced in writing this book was that the experts in probability theory tend to know very little algebraic number theory and *vice versa*: the experts in algebraic number theory do not really care much about probability theory. These two groups, “algebraists” and “probabilists,” are in fact very different kinds of mathematicians with totally different taste and different intuitions. It is hard to find a middle ground satisfying both groups, not to mention the readers who know little probability theory and little algebraic number theory. This forced us to include a lot of examples and “detours.”

The book grew from five partly-survey-partly-research papers of ours written between 1991 and 2000 (see [Be1, Be2, Be3, Be4, Be5]) and four more recent papers starting from 2010 (see [Be7, Be8, Be9, Be10]). In a nutshell, our work is a far-reaching extension of some classical results of Hardy–Littlewood and Ostrowski from the period of 1914–1920. In particular, we added the unifying “probabilistic viewpoint,” which is completely missing from the old papers. It is interesting to point out that for the generation of Hardy, number theory and probability sounded like a strange mismatch. Hardy once dismissively declared: “probability is not a notion of pure mathematics but of philosophy or physics” (Hardy made this statement before Kolmogorov’s axioms “legitimized” probability theory as a well-founded chapter in measure theory).

The main results of the book are Theorems 1.1, 1.2, 5.4, 5.6 (all about “randomness”) and the subject of “Area Principle versus superirregularity” (see, respectively, Proposition 1.18, Theorems 5.7 and 5.3, Sects. 5.4–5.10).

Since the two parts of the book are quite independent, the reader may start reading Part 1.3 first. We would recommend the reader to start with Sects. 1.1, 1.2, 5.1, and 5.2. An alternative way is to start with Sect. 2.5 and then go to Sects. 1.1, 1.2, 5.1, and 5.2.

The book is more or less self-contained. It should be readable to everybody with some basic knowledge of mathematics (second-year graduate students and up) who is interested in number theory and probability theory.

A few words about the notation. We constantly use the (rather standard) notation  $\{x\}$ ,  $\|x\|$ ,  $\lfloor x \rfloor$ ,  $\lceil x \rceil$ , which mean, in this order, the fractional part of a real number  $x$ , the distance of  $x$  from the nearest integer, and the lower and upper integral parts of  $x$  (for example,  $x = \{x\} + \lfloor x \rfloor$  and  $\|x\| = \min\{\{x\}, 1 - \{x\}\}$ ). A less well-known notation is

$$((x)) = \begin{cases} \{x\} - \frac{1}{2}, & \text{if } x \text{ is not an integer;} \\ 0, & \text{otherwise} \end{cases}$$

for the “sawtooth function,” which is permanently used in Part I of the book starting from Sect. 2.1. Throughout the letter  $c$  (or  $c_0, c_1, c_2, \dots$ ) denotes a generic constant, i.e., a positive constant that we could but do not care to determine. This constant may be absolute, or may depend upon the parameters involved in the theorem in question; it will not generally be the same constant. The well-known  $O$ -notation which occurs involves constants implicitly. It will generally be obvious on what, if any, parameters these constants depend. The natural (base  $e$ ) logarithm is denoted by  $\log$  (instead of  $\ln$  that we don’t use in the book). We use  $\log_2$  for the iterated logarithm, so  $\log_2 x = \log \log x$ ; we use  $\log x / \log 2$  to denote the binary (i.e., base 2) logarithm of  $x$ .

We are sure there are many errors in this first version of the book. We welcome any corrections, suggestions, and comments.

Piscataway, NJ, USA  
March 2014

József Beck





Probabilistic Diophantine Approximation  
Randomness in Lattice Point Counting

Beck, J.

2014, XVI, 487 p. 22 illus., Hardcover

ISBN: 978-3-319-10740-0