1

Introduction

1.1 Presentation

Different authors might define "probabilistic number theory" in different ways. Our point of view will be to see it as *the study of the asymptotic behavior of arithmetically defined sequences of probability measures, or random variables*. Thus the content of this book is based on examples of situations where we can say interesting things concerning such sequences. However, in Chapter 7, we will quickly survey some topics that might quite legitimately be seen as part of probabilistic number theory in a broader sense.

To illustrate what we have in mind, the most natural starting point is a famous result of Erdős and Kac.

Theorem 1.1.1 (the Erdős–Kac Theorem) *For any positive integer $n \geqslant 1$, let $\omega(n)$ denote the number of prime divisors of n, counted without multiplicity. Then, for any real numbers $a < b$, we have*

$$\lim_{N\to+\infty} \frac{1}{N}\left|\left\{1 \leqslant n \leqslant N \mid a \leqslant \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}} \leqslant b\right\}\right| = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx.$$

To spell out the connection between this statement and our slogan, one sequence of probability measures involved here is the sequence $(\mu_N)_{N\geqslant 1}$, defined as the uniform probability measure supported on the finite set $\Omega_N = \{1,\ldots,N\}$. This sequence is defined arithmetically, because the study of integers is part of arithmetic. The *asymptotic behavior* is revealed by the statement. Namely, consider the sequence of random variables

$$X_N(n) = \frac{\omega(n)-\log\log N}{\sqrt{\log\log N}}$$

defined on Ω_N for $N \geqslant 3$,[1] and the sequence (ν_N) of their probability distributions, which are (Borel) probability measures on $\mathbf{R}$ defined by

$$\nu_N(A) = \mu_N(X_N \in A) = \frac{1}{N}\left|\left\{1 \leqslant n \leqslant N \mid \frac{\omega(n) - \log\log N}{\sqrt{\log\log N}} \in A\right\}\right|$$

for any measurable set $A \subset \mathbf{R}$. These form another *arithmetically defined sequence of probability measures*, since primes are definitely arithmetic objects. Theorem 1.1.1 is, by basic probability theory, equivalent to the fact that the sequence (ν_N) converges in law to a standard Gaussian random variable as $N \rightarrow +\infty$. (We recall here that a sequence of real-valued random variables (X_N) converges in law to a random variable X if

$$\mathbf{E}(f(X_N)) \rightarrow \mathbf{E}(f(X))$$

for all bounded continuous functions $f : \mathbf{R} \rightarrow \mathbf{C}$, and that one can show that it is equivalent to

$$\mathbf{P}(a < X_N < b) \rightarrow \mathbf{P}(a < X < b)$$

for all $a < b$ such that $\mathbf{P}(X = a) = \mathbf{P}(X = b) = 0$; for the standard Gaussian, this means for all a and b; see Section B.3 for reminders about this.)

The Erdős–Kac Theorem is probably the simplest case where a natural deterministic arithmetic quantity (the number of prime factors of an integer), which is individually very hard to grasp, nevertheless exhibits a statistical or probabilistic behavior which fits a very common probability distribution. This is the prototype of the kinds of statements we will discuss (although sometimes the limiting measures will be far from standard!).

We will prove Theorem 1.1.1 in the next chapter. Before we do this, we will begin with a few results that are much more elementary but which may, with hindsight, be considered as the simplest cases of the type of results we want to describe.

1.2 How Does Probability Link with Number Theory Really?

Before embarking on this, however, it might be useful to give a rough idea of the way probability theory and arithmetic will combine to give interesting limit theorems like the Erdős–Kac Theorem. The strategy that we outline here

[1] Simply so that $\log\log N > 0$.

will be, in different guises, at the core of the strategy of the proofs of many theorems in this book.

We typically will be working with a sequence (X_n) of arithmetically interesting random variables, and we wish to prove that it converges in law. In many cases, we do this with a two-step process.

(1) We begin by approximating (X_n) by another sequence (Y_n), in such a way that convergence in law of these approximations implies that of (X_n), with the same limit. In other words, we see Y_n as a kind of perturbation of X_n, which is small enough to preserve convergence in law. Notably, the approximation might be of different sorts: the difference $\mathsf{X}_n - \mathsf{Y}_n$ might, for instance, converge to 0 in probability, or in some L^p-space; in fact, we will sometimes encounter a process of successive approximations, where the successive perturbations are small in different senses, before reaching a convenient approximation Y_n (this is the case in the proof of Theorem 4.1.2).

(2) Having found a good approximation Y_n, we prove that it converges in law using a probabilistic criterion that is sufficiently robust to apply; typical examples are the method of moments, and the convergence theorem of P. Lévy based on characteristic functions (i.e., Fourier transforms), because analytic number theory often gives tools to compute approximately such invariants of arithmetically defined random variables.

Both steps are sometimes quite easy to motivate using some heuristic arguments (for instance, when X_n or Y_n are represented as a sum of various terms, we might guess that these are “approximately independent,” to lead to a limit similar to that of sums of independent random variables), but they may also involve quite subtle ideas.

We will not dwell further on this overarching strategy, but the reader will be able to recognize how it fits into this skeleton when we discuss the steps of the proof of some of the main theorems.

In many papers written by (or for) analytic number theorists, the approximations of Step 1, as well as (say) the moment computations of Step 2, are performed using notation, terminology and normalizations coming from the customs and standards of analytic number theory. In this book, we will try to express them instead, as much as possible, in good probabilistic style (e.g., we attempt to mention as little as possible the “elementary events” of the underlying probability space). This is usually simply a matter of cosmetic transformations, but sometimes it leads to slightly different emphasis, in particular concerning the nature of the approximations in Step 1. We suggest that the reader compare our presentation with that of some of the original source papers, in order to assess whether this style is enlightening (as we often find it to be), or not.

1.3 A Prototype: Integers in Arithmetic Progressions

As mentioned above, we begin with a result that is so elementary that it is usually not presented as a separate statement (let alone as a theorem!). Nevertheless, as we will see, it is the basic ingredient (and explanation) for the Erdős–Kac Theorem, and generalizations of it become quite quickly very deep.

Theorem 1.3.1 *For* $\mathrm{N} \geqslant 1$, *let* $\Omega_{\mathrm{N}} = \{1, \ldots, \mathrm{N}\}$ *with the uniform probability measure* $\mathbf{P}_{\mathrm{N}}$. *Fix an integer* $q \geqslant 1$, *and denote by* $\pi_q : \mathbf{Z} \longrightarrow \mathbf{Z}/q\mathbf{Z}$ *the reduction modulo* q *map. Let* X_{N} *be the random variables given by* $\mathsf{X}_{\mathrm{N}}(n) = \pi_q(n)$ *for* $n \in \Omega_{\mathrm{N}}$.

As $\mathrm{N} \to +\infty$, *the random variables* X_{N} *converge in law to the uniform probability measure* μ_q *on* $\mathbf{Z}/q\mathbf{Z}$. *In fact, for any function*

$$f : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{C},$$

we have

$$\big|\mathbf{E}(f(\mathsf{X}_{\mathrm{N}})) - \mathbf{E}(f)\big| \leqslant \frac{2}{\mathrm{N}} \|f\|_1, \tag{1.1}$$

where

$$\|f\|_1 = \sum_{a \in \mathbf{Z}/q\mathbf{Z}} |f(a)|.$$

Proof It is enough to prove (1.1), which gives the convergence in law by letting $\mathrm{N} \to +\infty$. This is quite simple. By definition, we have

$$\mathbf{E}(f(\mathsf{X}_{\mathrm{N}})) = \frac{1}{\mathrm{N}} \sum_{1 \leqslant n \leqslant \mathrm{N}} f(\pi_q(n))$$

and

$$\mathbf{E}(f) = \frac{1}{q} \sum_{a \in \mathbf{Z}/q\mathbf{Z}} f(a).$$

The idea is then clear: among the integers $1 \leqslant n \leqslant \mathrm{N}$, roughly N/q are in any given residue class $a \pmod q$, and if we use this approximation in the first formula, we obtain precisely the second.

To do this in detail, we gather the integers in the sum according to their residue class a modulo q. This gives

$$\frac{1}{N}\sum_{1\leqslant n\leqslant N} f(\pi_q(n)) = \sum_{a\in \mathbf{Z}/q\mathbf{Z}} f(a)\times\frac{1}{N}\sum_{\substack{1\leqslant n\leqslant N\\ n\equiv a\,(\mathrm{mod}\,q)}} 1.$$

The inner sum, for each a, counts the number of integers n in the interval $1\leqslant n\leqslant N$ such that the remainder under division by q is a. These integers n can be written $n = mq + a$ for some $m \in \mathbf{Z}$, if we view a as an actual integer, and therefore it is enough to count those integers $m \in \mathbf{Z}$ for which $1 \leqslant mq + a \leqslant N$. The condition translates to

$$\frac{1-a}{q}\leqslant m\leqslant\frac{N-a}{q},$$

and therefore we are reduced to counting integers *in an interval*. This is not difficult, although we have to be careful with boundary terms, since the bounds of the interval are not necessarily integers. The length of the interval is $(N-a)/q-(1-a)/q = (N-1)/q$. In general, in an interval $[\alpha,\beta]$ with $\alpha\leqslant\beta$, the number $N_{\alpha,\beta}$ of integers satisfies

$$\beta-\alpha-1\leqslant N_{\alpha,\beta}\leqslant\beta-\alpha+1$$

(and the boundary contributions should not be forgotten, although they are typically negligible when the interval is long enough).

Hence the number N_a of values of m satisfies

$$\frac{N-1}{q}-1\leqslant N_a\leqslant\frac{N-1}{q}+1, \tag{1.2}$$

and therefore

$$\left|N_a-\frac{N}{q}\right|\leqslant 1+\frac{1}{q}.$$

By summing over a in $\mathbf{Z}/q\mathbf{Z}$, we deduce now that

$$\begin{aligned}\left|\frac{1}{N}\sum_{1\leqslant n\leqslant N} f(\pi_q(n)) - \frac{1}{q}\sum_{a\in\mathbf{Z}/q\mathbf{Z}} f(a)\right| &= \left|\sum_{a\in\mathbf{Z}/q\mathbf{Z}} f(a)\left(\frac{N_a}{N}-\frac{1}{q}\right)\right|\\ &\leqslant \frac{1+q^{-1}}{N}\sum_{a\in\mathbf{Z}/q\mathbf{Z}}|f(a)|\leqslant\frac{2}{N}\|f\|_1.\end{aligned}$$

□

Remark 1.3.2 As a matter of notation, we will sometimes remove the variable N from the notation of random variables, since the value of N is usually made clear by the context, frequently because of its appearance in an expression involving $\mathbf{P}_{\mathrm{N}}(\cdot)$ or $\mathbf{E}_{\mathrm{N}}(\cdot)$, which refers to the probability and expectation on Ω_{N}.

Despite its simplicity, this result already brings up a number of important features that will occur extensively in later chapters.

A first remark is that we actually proved something much stronger than the statement of convergence in law: the bound (1.1) gives a rather precise estimate of the speed of convergence of expectations (or probabilities) computed using the law of X_{N} to those computed using the limit uniform distribution μ_q. Most importantly, as we will see shortly, these estimates are uniform in terms of q, and give us information on convergence, or more properly speaking on the "distance" between the law of X_{N} and μ_q, even if q depends on N in some way.

To be more precise, take f to be the characteristic function of a residue class $a \in \mathbf{Z}/q\mathbf{Z}$. Then since $\mathbf{E}(f) = 1/q$, we get

$$\left|\mathbf{P}(\pi_q(n) = a) - \frac{1}{q}\right| \leqslant \frac{2}{\mathrm{N}}.$$

This is nontrivial information as long as q is a bit smaller than N. Thus, this states that the probability that $n \leqslant \mathrm{N}$ is congruent to a modulo q is close to the intuitive probability $1/q$, uniformly for all q just a bit smaller than N, and also uniformly for all residue classes. We will see, both below and in many similar situations, that uniformity aspects are essential in applications.

The second remark concerns the interpretation of the result. Theorem 1.3.1 can explain what is meant by such intuitive statements as *the probability that an integer is divisible by* 2 *is* 1/2. Namely, this is the probability, according to the uniform measure on $\mathbf{Z}/2\mathbf{Z}$, of the set $\{0\}$, and this is simply the limit given by the convergence in law of the variables $\pi_2(n)$ defined on Ω_{N} to the uniform measure μ_2.

This idea applies to many other similar-sounding problems. The most elementary among these can often be solved using Theorem 1.3.1. We present one famous example: what is the "probability" that an integer $n \geqslant 1$ is squarefree, which means that n is *not* divisible by a square m^2 for some integer $m \geqslant 2$ (or, equivalently, by the square of some prime number)? Here the interpretation is that this probability should be

$$\lim_{\mathrm{N}\to+\infty} \frac{1}{\mathrm{N}} |\{1 \leqslant n \leqslant \mathrm{N} \mid n \text{ is squarefree}\}|.$$

If we prefer (as we do) to speak of sequences of random variables, we can take the sequence of Bernoulli random variables B_N indicators of the event that $n \in \Omega_N$ is squarefree, so that

$$\mathbf{P}(\mathsf{B}_N = 1) = \frac{1}{N}|\{1 \leqslant n \leqslant N \mid \; n \text{ is squarefree}\}|.$$

We then ask about the limit in law of (B_N). The answer is as follows:

Proposition 1.3.3 *The sequence* (B_N) *converges in law to a Bernoulli random variable* B *with* $\mathbf{P}(B = 1) = \frac{6}{\pi^2}$. *In other words, the "probability" that an integer n is squarefree, in the interpretation discussed above, is* $6/\pi^2$.

Proof The idea is to use inclusion-exclusion: to say that n is squarefree means that it is not divisible by the square p^2 of any prime number. Thus, if we denote by $\mathbf{P}_N$ the probability measure on Ω_N, we have

$$\mathbf{P}_N(n \text{ is squarefree}) = \mathbf{P}_N\Big(\bigcap_{p \text{ prime}} \{p^2 \text{ does not divide } n\}\Big).$$

There is one key step now that is both obvious and crucial: because of the nature of Ω_N, the infinite intersection may be replaced by the intersection over primes $p \leqslant \sqrt{N}$, since all integers in Ω_N are $\leqslant N$. Applying the inclusion-exclusion formula, we obtain

$$\mathbf{P}_N\Big(\bigcap_{p\leqslant N^{1/2}} \{p^2 \text{ does not divide } n\}\Big) = \sum_{I} (-1)^{|I|}\, \mathbf{P}_N\Big(\bigcap_{p\in I}\{p^2 \text{ divides } n\}\Big), \tag{1.3}$$

where I runs over the set of subsets of the set $\{p \leqslant N^{1/2}\}$ of primes $\leqslant N^{1/2}$, and $|I|$ is the cardinality of I. But, by the Chinese Remainder Theorem, we have

$$\bigcap_{p\in I}\{p^2 \text{ divides } n\} = \{d_I^2 \text{ divides } n\},$$

where d_I is the product of the primes in I. Once more, note that this set is empty if $d_I^2 > N$. Moreover, the fundamental theorem of arithmetic shows that $I \mapsto d_I$ is injective, and we can recover $|I|$ also from d_I as the number of prime factors of d_I. Therefore, we get

$$\mathbf{P}_N(n \text{ is squarefree}) = \sum_{d\leqslant N^{1/2}} \mu(d)\, \mathbf{P}_N(d^2 \text{ divides } n),$$

where $\mu(d)$ is the Möbius function, defined for integers $d \geqslant 1$ by

$$\mu(d) = \begin{cases} 0 & \text{if } d \text{ is not squarefree,} \\ (-1)^k & \text{if } d = p_1 \cdots p_k \text{ with } p_i \text{ distinct primes} \end{cases}$$

(see Definition C.1.3).

But d^2 divides n if and only if the image of n by reduction modulo d^2 is 0. By Theorem 1.3.1 applied with $q = d^2$ for all $d \leqslant \mathrm{N}^{1/2}$, with f the indicator function of the residue class of 0, we get

$$\mathbf{P}_{\mathrm{N}}(d^2 \text{ divides } n) = \frac{1}{d^2} + \mathrm{O}(\mathrm{N}^{-1})$$

for all d, where the implied constant in the $\mathrm{O}(\cdot)$ symbol is independent of d (in fact, it is at most 2). Note in passing how we use crucially here the fact that Theorem 1.3.1 was uniform and explicit with respect to the parameter q.

Summing the last formula over $d \leqslant \mathrm{N}^{1/2}$, we deduce

$$\mathbf{P}_{\mathrm{N}}(n \text{ is squarefree}) = \sum_{d \leqslant n^{1/2}} \frac{\mu(d)}{d^2} + \mathrm{O}\left(\frac{1}{\sqrt{\mathrm{N}}}\right).$$

Since the series with terms $1/d^2$ converges, this shows the existence of the limit, and that $(\mathsf{B}_{\mathrm{N}})$ converges in law as $\mathrm{N} \to +\infty$ to a Bernoulli random variable B with success probability

$$\mathbf{P}(\mathrm{B} = 1) = \sum_{d \geqslant 1} \frac{\mu(d)}{d^2}, \qquad \mathbf{P}(\mathrm{B} = 0) = 1 - \sum_{d \geqslant 1} \frac{\mu(d)}{d^2}.$$

It is a well-known fact (the “Basel problem,” first solved by Euler; see Exercise 1.3.4 for a proof) that

$$\sum_{d \geqslant 1} \frac{1}{d^2} = \frac{\pi^2}{6}.$$

Moreover, a basic property of the Möbius function states that

$$\sum_{d \geqslant 1} \frac{\mu(d)}{d^s} = \frac{1}{\zeta(s)}$$

for any complex number s with $\mathrm{Re}(s) > 1$, where

$$\zeta(s) = \sum_{d \geqslant 1} \frac{1}{d^s}$$

is the Riemann zeta function (Corollary C.1.5), and hence we get

$$\sum_{d\geqslant 1}\frac{\mu(d)}{d^2}=\frac{1}{\zeta(2)}=\frac{6}{\pi^2}. \qquad \square$$

Exercise 1.3.4 In this exercise, we explain a proof of Euler's formula $\zeta(2)=\pi^2/6$.

(1) Assuming that

$$\frac{\sin(\pi x)}{\pi x}=\prod_{n\geqslant 1}\Bigl(1-\frac{x^2}{n^2}\Bigr)$$

(another formula of Euler), find a heuristic proof of $\zeta(2)=\pi^2/6$. [**Hint**: First, express the sum of the inverses of the roots of a polynomial (with nonzero constant term) in terms of its coefficients.]

The following argument, due to Cauchy, can be seen as a way to make rigorous the previous idea.

(2) Show that for $n\geqslant 1$ and $x\in\mathbf{R}-\pi\mathbf{Z}$, we have

$$\frac{\sin(nx)}{(\sin x)^n}=\sum_{0\leqslant m\leqslant n/2}(-1)^m\binom{n}{2m+1}\operatorname{cotan}(x)^{n-(2m+1)}.$$

(3) Let $m\geqslant 1$ be an integer, and let $n=2m+1$. Show that

$$\sum_{r=1}^{m}\operatorname{cotan}\Bigl(\frac{r\pi}{n}\Bigr)^2=\frac{2m(2m-1)}{6}$$

and

$$\sum_{r=1}^{m}\frac{1}{\sin\Bigl(\frac{r\pi}{n}\Bigr)^2}=\frac{2m(2m+2)}{6}.$$

[**Hint**: Using (1), view the numbers $\operatorname{cotan}(r\pi/n)^2$ as the roots of a polynomial of degree m, and use the formula for the sum of the roots of a polynomial.]

(4) Deduce that

$$\frac{2m(2m-1)}{6}<\sum_{k=1}^{m}\Bigl(\frac{2m+1}{k\pi}\Bigr)^2<\frac{2m(2m+2)}{6},$$

and conclude.

The proof of Proposition 1.3.3 above was written in probabilistic style, emphasizing the connection with Theorem 1.3.1. It can be expressed more straightforwardly as a sequence of manipulation with finite sums, using the formula

$$\sum_{d^2\mid n}\mu(d)=\begin{cases}1 & \text{if } n \text{ is squarefree,}\\ 0 & \text{otherwise}\end{cases} \tag{1.4}$$

for $n \geqslant 1$ (which is implicit in our discussion) and the approximation

$$\sum_{\substack{1\leqslant n\leqslant \mathrm{N}\\ d\mid n}} 1=\frac{\mathrm{N}}{d}+\mathrm{O}(1)$$

for the number of integers in an interval which are divisible by some $d \geqslant 1$. This goes as follows:

$$\begin{aligned}\sum_{\substack{n\leqslant \mathrm{N}\\ n \text{ squarefree}}} 1=\sum_{n\leqslant \mathrm{N}}\sum_{d^2\mid n}\mu(d) &= \sum_{d\leqslant\sqrt{\mathrm{N}}}\mu(d)\sum_{\substack{n\leqslant \mathrm{N}\\ d^2\mid n}} 1\\ &=\sum_{d\leqslant\sqrt{\mathrm{N}}}\mu(d)\left(\frac{\mathrm{N}}{d^2}+\mathrm{O}(1)\right)\\ &=\mathrm{N}\sum_{d}\frac{\mu(d)}{d^2}+\mathrm{O}(\sqrt{\mathrm{N}}).\end{aligned}$$

Obviously, this is much shorter, although one needs to know the formula (1.4), which was implicitly derived in the previous proof.[2] But there is something quite important to be gained from the probabilistic viewpoint, which might be missed by reading too quickly the second proof. Indeed, in formulas like (1.3) (or many others), the precise nature of the underlying probability space Ω_{N} is quite hidden – as is customary in probability where this is often not really relevant. In our situation, this suggests naturally to study similar problems for *different* sequences of integer-valued random variables rather than taking integers uniformly between 1 and N.

This has indeed been done, and in many different ways. But even before looking at any example, we can predict that some new – interesting – phenomena will arise when doing so. Indeed, even if our first proof of Proposition 1.3.3 was written in a very general probabilistic language, it did use one special feature of Ω_{N}: it only contains integers $n \leqslant \mathrm{N}$, and even more particularly, it does not contain any element divisible by d^2 for d larger than $\sqrt{\mathrm{N}}$. (More probabilistically, the probability $\mathbf{P}_{\mathrm{N}}(d^2 \text{ divides } n)$ is then zero.)

[2] Readers who are already well versed in analytic number theory might find it useful to translate back and forth various estimates written in probabilistic style in this book.