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 $\Omega_N = \{1, \dots, N\}$. 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

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

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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 $N \to +\infty$. (We recall here that a sequence of real-valued random variables (X_N) converges in law to a random variable X if

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

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

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

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 $\log\log N > 0$.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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 $X_n - Y_n$ 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* $N \geqslant 1$, *let* $\Omega_N = \{1, \dots, N\}$ *with the uniform probability measure* $\mathbf{P}_N$. *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* $\mathsf{X}_N(n) = \pi_q(n)$ *for* $n \in \Omega_N$.

As $N \to +\infty$, *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

$$\left|\mathbf{E}(f(\mathsf{X}_N)) - \mathbf{E}(f)\right| \leqslant \frac{2}{N}\|f\|_1, \tag{1.1}$$

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 $N \to +\infty$. This is quite simple. By definition, we have

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

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 $1 \leqslant n \leqslant N$, 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.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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

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

The inner sum, for each a, counts the number of integers n in the interval $1 \leqslant n \leqslant \mathrm{N}$ 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 $1 \leqslant mq + a \leqslant \mathrm{N}$. The condition translates to

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

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 $(\mathrm{N}-a)/q-(1-a)/q = (\mathrm{N}-1)/q$. In general, in an interval $[\alpha, \beta]$ with $\alpha \leqslant \beta$, the number $\mathrm{N}_{\alpha,\beta}$ of integers satisfies

$$\beta - \alpha - 1 \leqslant \mathrm{N}_{\alpha,\beta} \leqslant \beta - \alpha + 1$$

(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

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

and therefore

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

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

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

□

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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

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

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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

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

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

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

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

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

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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

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

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

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

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

□

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

(1) Assuming that

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

(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

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

and

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

[**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

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

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

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

straightforwardly as a sequence of manipulation with finite sums, using the formula

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

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

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

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

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

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.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

Now consider the following extension of the problem, which is certainly one of the first that may come to mind beyond our initial setting: we fix an irreducible polynomial $P \in \mathbf{Z}[X]$ of degree $m \geqslant 1$, and consider new Bernoulli random variables $\mathsf{B}_{P,N}$ which are indicators of the event that $P(n)$ is squarefree on Ω_N (instead of n itself). Asking about the limit of these random variables means asking for the "probability" that $P(n)$ is squarefree, when $1 \leqslant n \leqslant N$. But although there is an elementary analogue of Theorem 1.3.1, it is easy to see that this does *not* give enough control of

$$\mathbf{P}_N(d^2 \text{ divides } P(n))$$

when d is "too large" compared with N. And this explains partly why, in fact, as of 2020 at least, *there is not even a* single *irreducible polynomial* $P \in \mathbf{Z}[X]$ *of degree* 4 *or higher for which it is known that* $P(n)$ *is squarefree infinitely often.*

Exercise 1.3.5 (1) Let $k \geqslant 2$ be an integer. Compute the "probability," in the same sense as in Proposition 1.3.3, that an integer n is k-free, that is, that there is no integer $m \geqslant 2$ such that m^k divides n.

(2) Compute the "probability" that two integers n_1 and n_2 are coprime, in the sense of taking the corresponding Bernoulli random variables on $\Omega_N \times \Omega_N$ and their limit as $N \to +\infty$.

Exercise 1.3.6 Let $P \in \mathbf{Z}[X]$ be an irreducible polynomial of degree $m \geqslant 1$. For $q \geqslant 1$, let π_q be the projection from $\mathbf{Z}$ to $\mathbf{Z}/q\mathbf{Z}$ as before.

(1) Show that for any $q \geqslant 1$, the random variables $\mathsf{X}_N(n) = \pi_q(P(n))$ converge in law to a probability measure $\mu_{P,q}$ on $\mathbf{Z}/q\mathbf{Z}$. Is $\mu_{P,q}$ uniform?

(2) Find values of T, depending on N and as large as possible, such that

$$\mathbf{P}_N(P(n) \text{ is not divisible by } p^2 \text{ for } p \leqslant T) > 0.$$

How large should T be so that this implies straightforwardly that

$$\{n \geqslant 1 \mid P(n) \text{ is squarefree}\}$$

is infinite?

(3) Prove that the set

$$\{n \geqslant 1 \mid P(n) \text{ is } (m+1)\text{-free}\}$$

is infinite.

We conclude this section with another very important feature of Theorem 1.3.1 from the probabilistic point of view, namely, its link with independence.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

If q_1 and q_2 are positive integers which are coprime, then the Chinese Remainder Theorem implies that the map

$$\begin{cases} \mathbf{Z}/q_1q_2\mathbf{Z} \longrightarrow \mathbf{Z}/q_1\mathbf{Z} \times \mathbf{Z}/q_2\mathbf{Z}, \\ x \mapsto (x \pmod{q_1}, x \pmod{q_2}) \end{cases}$$

is a bijection (in fact, a ring isomorphism). Under this bijection, the uniform probability measure $\mu_{q_1q_2}$ on $\mathbf{Z}/q_1q_2\mathbf{Z}$ corresponds to the product measure $\mu_{q_1} \otimes \mu_{q_2}$. In particular, the random variables $x \mapsto x \pmod{q_1}$ and $x \mapsto x \pmod{q_2}$ on $\mathbf{Z}/q_1q_2\mathbf{Z}$ are independent.

The interpretation of this is that the random variables π_{q_1} and π_{q_2} on Ω_{N} are *asymptotically independent* as $\mathrm{N} \to +\infty$, in the sense that

$$\lim_{\mathrm{N}\to+\infty} \mathbf{P}_{\mathrm{N}}(\pi_{q_1}(n) = a \text{ and } \pi_{q_2}(n) = b) = \frac{1}{q_1q_2} = \left(\lim_{\mathrm{N}\to+\infty} \mathbf{P}_{\mathrm{N}}(\pi_{q_1}(n) = a)\right) \times \left(\lim_{\mathrm{N}\to+\infty} \mathbf{P}_{\mathrm{N}}(\pi_{q_2}(n) = b)\right)$$

for all $(a, b) \in \mathbf{Z}^2$. Intuitively, one would say that *divisibility by q_1 and q_2 are independent*, and especially that *divisibility by distinct primes are independent events*. We summarize this in the following extremely useful proposition:

Proposition 1.3.7 *For $\mathrm{N} \geqslant 1$, let $\Omega_{\mathrm{N}} = \{1, \dots, \mathrm{N}\}$ with the uniform probability measure $\mathbf{P}_{\mathrm{N}}$. Fix a finite set S of pairwise coprime integers.*

As $\mathrm{N} \to +\infty$, the vector $(\pi_q)_{q\in\mathrm{S}}$ seen as random vector on Ω_{N} with values in

$$\mathrm{X}_{\mathrm{S}} = \prod_{q\in\mathrm{S}} \mathbf{Z}/q\mathbf{Z}$$

converges in law to a vector of independent and uniform random variables. In fact, for any function

$$f : \mathrm{X}_{\mathrm{S}} \longrightarrow \mathbf{C},$$

we have

$$\left|\mathbf{E}(f((\pi_q)_{q\in\mathrm{S}})) - \mathbf{E}(f)\right| \leqslant \frac{2}{\mathrm{N}}\|f\|_1. \tag{1.5}$$

Proof This is just an elaboration of the previous discussion. Let r be the product of the elements of S. Then the Chinese Remainder Theorem gives a ring-isomorphism $\mathrm{X}_{\mathrm{S}} \longrightarrow \mathbf{Z}/r\mathbf{Z}$ such that the uniform measure μ_r on the right-hand side corresponds to the product of the uniform measures on X_{S}.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

Thus f can be identified with a function $g\colon \mathbf{Z}/r\mathbf{Z} \longrightarrow \mathbf{C}$, and its expectation to the expectation of g according to μ_r. By Theorem 1.3.1, we get

$$\bigl|\mathbf{E}(f((\pi_q)_{q\in S})) - \mathbf{E}(f)\bigr| = \bigl|\mathbf{E}(g(\pi_r)) - \mathbf{E}(g)\bigr| \leqslant \frac{2\|g\|_1}{N},$$

which is the desired result since f and g have also the same ℓ^1 norm. □

Remark 1.3.8 (1) Note that the random variables obtained by reduction modulo two coprime integers are not exactly independent: it is not true in general that

$$\mathbf{P}_N(\pi_{q_1}(n) = a \text{ and } \pi_{q_2}(n) = b) = \mathbf{P}_N(\pi_{q_1}(n) = a)\,\mathbf{P}_N(\pi_{q_2}(n) = b).$$

This is the source of many interesting aspects of probabilistic number theory where classical ideas and concepts of probability for sequences of independent random variables are generalized or "tested" in a context where independence only holds in an asymptotic or approximate sense.

(2) There is one subtle point that appears in quantitative applications of Theorem 1.3.1 and Proposition 1.3.7 that is worth mentioning. Given an integer $q \geqslant 1$, certain functions f on $\mathbf{Z}/q\mathbf{Z}$ might have a large norm $\|f\|_1$, and yet they may have expressions as linear combinations of functions $\widetilde{f}$ on certain spaces $\mathbf{Z}/d\mathbf{Z}$, where d is a divisor of q, which have much smaller norms $\|\widetilde{f}\|_1$. Taking such possibilities into account and arguing modulo d instead of modulo q may lead to stronger estimates for the error

$$\mathbf{E}_N(f(\pi_q(n))) - \mathbf{E}(f)$$

than those we have written down in terms of $\|f\|_1$. This is, for instance, especially clear if we take f to be a nonzero constant, in which case the difference is actually 0, but $\|f\|_1$ is of size q.

One can incorporate formally these improvements by using a different norm than $\|f\|_1$, as we now explain.

Let $q \geqslant 1$ be an integer. Let Φ_q be the set of functions $\varphi_{d,a}\colon \mathbf{Z}/q\mathbf{Z} \to \mathbf{C}$ which are characteristic functions of classes $x \equiv a \pmod d$ for some positive divisor $d \mid q$ and some $a \in \mathbf{Z}/d\mathbf{Z}$ (these are well-defined functions modulo q). In particular, the function $\varphi_{q,a}$ is just the delta function at a in $\mathbf{Z}/q\mathbf{Z}$, and $\varphi_{1,0}$ is the constant function 1.

For an arbitrary function $f\colon \mathbf{Z}/q\mathbf{Z} \to \mathbf{C}$, let

$$\|f\|_{c,1} = \inf\left\{ \sum_{d\mid q} \sum_{a \pmod d} |\lambda_{d,a}| \;\mid\; f = \sum_{d\mid q} \sum_{a \pmod d} \lambda_{d,a}\varphi_{d,a} \right\}.$$

This defines a norm on the space of functions on $\mathbf{Z}/q\mathbf{Z}$ (the subscript c refers to *congruences*); the norm $\|f\|_{c,1}$ measures how simply the function f may be

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

expressed as a linear combination of indicator functions of congruence classes modulo divisors of q.[3] Note that $\|f\|_{c,1} \leqslant \|f\|_1$, because one always has the representation

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

Now the estimates (1.1) and (1.5) can be improved to

$$\left|\mathbf{E}(f(\mathsf{X}_\mathrm{N})) - \mathbf{E}(f)\right| \leqslant \frac{2}{\mathrm{N}}\|f\|_{c,1}, \tag{1.6}$$

$$\left|\mathbf{E}(f((\pi_q)_{q\in\mathrm{S}})) - \mathbf{E}(f)\right| \leqslant \frac{2}{\mathrm{N}}\|f\|_{c,1}, \tag{1.7}$$

respectively. Indeed, it suffices (using linearity and the triangle inequality) to check this for $f = \varphi_{d,a}$ for some divisor $d \mid q$ and some $a \in \mathbf{Z}/d\mathbf{Z}$ (with $\|\varphi_{d,a}\|_{c,1}$ replaced by 1 in the right-hand side), in which case the difference (in the first case) is

$$\frac{1}{\mathrm{N}} \sum_{\substack{n \leqslant \mathrm{N} \\ n \equiv a \pmod d}} 1 - \frac{1}{q} \sum_{\substack{x \pmod q \\ x \equiv a \pmod d}} 1 = \frac{1}{\mathrm{N}} \sum_{\substack{n \leqslant \mathrm{N} \\ n \equiv a \pmod d}} 1 - \frac{1}{d},$$

which reduces to the case of single element modulo d, for which we now apply Theorem 1.3.1.

Another corollary of these elementary statements identifies the limiting distribution of the valuations of integers. To state it, we denote by S_N the identity random variable on the probability space $\Omega_\mathrm{N} = \{1, \dots, \mathrm{N}\}$ with uniform probability measure of Theorem 1.3.1.

Corollary 1.3.9 *For p prime, let v_p denote the p-adic valuation on $\mathbf{Z}$. The random vectors $(v_p(\mathsf{S}_\mathrm{N}))_p$ converge in law, in the sense of finite distributions, to a sequence $(\mathrm{V}_p)_p$ of independent geometric random variables with*

$$\mathbf{P}(\mathrm{V}_p = k) = \left(1 - \frac{1}{p}\right)\frac{1}{p^k}$$

for $k \geqslant 0$. In other words, for any finite set of primes S and any nonnegative integers $(k_p)_{p\in\mathrm{S}}$, we have

$$\lim_{\mathrm{N}\to+\infty} \mathbf{P}_\mathrm{N}(v_p(\mathsf{S}_\mathrm{N}) = k_p \text{ for } p \in \mathrm{S}) = \prod_{p\in\mathrm{S}} \mathbf{P}(\mathrm{V}_p = k_p).$$

[3] In terms of functional analysis, this means that this is a quotient norm of the ℓ^1 norm on the space with basis Φ_q.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

Proof For a given prime p and integer $k \geqslant 0$, the condition that $v_p(n) = k$ means that $n \pmod{p^{k+1}}$ belongs to the subset in $\mathbf{Z}/p^{k+1}\mathbf{Z}$ of residue classes of the form bp^k where $1 \leqslant b \leqslant p-1$; by Theorem 1.3.1, we therefore have

$$\lim_{N\to+\infty} \mathbf{P}_N(v_p(S_N) = k) = \frac{p-1}{p^{k+1}} = \mathbf{P}(V_p = k).$$

Proposition 1.3.7 then shows that this extends to any finite set of primes. □

Example 1.3.10 Getting quantitative estimates in this corollary is a good example of Remark 1.3.8 (2). We illustrate this point in the simplest case, which will be used in Section 2.2.

Consider two primes $p \neq q$ and the probability

$$\mathbf{P}_N(v_p(S_N) = v_q(S_N) = 1).$$

The indicator function φ of this event is naturally defined modulo p^2q^2, and its norm $\|\varphi\|_1$ is the number of integers modulo p^2q^2 that are multiples of pq, but not of p^2 or q^2. By inclusion-exclusion, this means that $\|\varphi\|_1 = (p-1)(q-1)$. On the other hand, we have $\varphi = \varphi_1 - \varphi_2 - \varphi_3 + \varphi_4$ where

- the function φ_1 is defined modulo pq as the indicator of the class 0;
- the function φ_2 is defined modulo p^2q as the indicator of the class 0;
- the function φ_3 is defined modulo pq^2 as the indicator of the class 0;
- the function φ_4 is defined modulo p^2q^2 as the indicator of the class 0.

Hence, in the notation of Remark 1.3.8 (2), we have $\|\varphi\|_{c,1} \leqslant 4$; using this remark, or by applying Theorem 1.3.1 four times, we get

$$\mathbf{P}_N(v_p(S_N) = v_q(S_N) = 1) = \frac{1}{pq}\left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right) + O\left(\frac{1}{N}\right),$$

instead of having an error term of size pq/N, as suggested by a direct application of (1.1).

1.4 Another Prototype: The Distribution of the Euler Function

Although Proposition 1.3.7 is extremely simple, it is the only necessary arithmetic ingredient in the proof of a result that is another prototype of probabilistic number theory in our sense. This is a theorem proved by Schoenberg [108] in 1928, which therefore predates the Erdős–Kac Theorem by about ten years (although Schoenberg phrased the result quite differently, since this date is also before Kolmogorov's formalization of probability theory).

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

The Euler "totient" function is defined for integers $n \geqslant 1$ by $\varphi(n) = |(\mathbf{Z}/n\mathbf{Z})^\times|$ (the number of invertible residue classes modulo n). By the Chinese Remainder Theorem (see Example C.1.8), this function is multiplicative, in the sense that $\varphi(n_1 n_2) = \varphi(n_1)\varphi(n_2)$ for n_1 coprime to n_2. Computing $\varphi(p^k) = p^k - p^{k-1} = p^k(1 - 1/p)$ for p prime and $k \geqslant 1$, one deduces that

$$\frac{\varphi(n)}{n} = \prod_{p|n}\left(1 - \frac{1}{p}\right)$$

for all integers $n \geqslant 1$ (where the product is over primes p dividing n).

Now define random variables F_N on $\Omega_\mathrm{N} = \{1, \ldots, \mathrm{N}\}$ (with the uniform probability measure as before) by

$$\mathsf{F}_\mathrm{N}(n) = \frac{\varphi(n)}{n}.$$

We will prove that the sequence $(\mathsf{F}_\mathrm{N})_{\mathrm{N}\geqslant 1}$ converges in law, and identify its limiting distribution. For this purpose, let $(\mathrm{B}_p)_p$ be a sequence of independent Bernoulli random variables, indexed by primes, with

$$\mathbf{P}(\mathrm{B}_p = 1) = \frac{1}{p} \quad \text{and} \quad \mathbf{P}(\mathrm{B}_p = 0) = 1 - \frac{1}{p}$$

(such random variables will also occur prominently in the next chapter).

Proposition 1.4.1 *The random variables F_N converge in law to the random variable given by*

$$\mathrm{F} = \prod_p \left(1 - \frac{\mathrm{B}_p}{p}\right),$$

where the infinite product ranges over all primes and converges almost surely.

This proposition is not only a good illustration of limiting behavior of arithmetic random variables, but the proof that we give, which emphasizes probabilistic methods, is an excellent introduction to a number of techniques that will occur later in more complicated contexts. Before we begin, note how the limiting random variable is highly nongeneric, and in fact retains some arithmetic information, since it is a product over primes. In particular, although the arithmetic content does not go beyond Proposition 1.3.7, this theorem is certainly not an obvious fact.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

Proof For $M \geqslant 1$, we denote by $\mathsf{F}_{N,M}$ the random variable on Ω_N defined by

$$\mathsf{F}_{N,M}(n) = \prod_{\substack{p|n \\ p\leqslant M}} \left(1 - \frac{1}{p}\right).$$

It is natural to think of these as approximations to F_N. On the other hand, for a fixed M, these are finite products and hence easier to handle. We will use a fairly simple "perturbation lemma" to prove the convergence in law of the sequence $(\mathsf{F}_N)_{N\geqslant 1}$ from the understanding of the behavior of $\mathsf{F}_{N,M}$. The lemma is precisely Proposition B.4.4, which the reader should read now.[4]

First, we fix $M \geqslant 1$. Since only primes $p \leqslant M$ occur in the definition of $\mathsf{F}_{N,M}$, it follows from Proposition 1.3.7 that the random variables $\mathsf{F}_{N,M}$ converge in law as $N \to +\infty$ to the random variable

$$F_M = \prod_{p\leqslant M} \left(1 - \frac{B_p}{p}\right).$$

Thus Assumption (1) in Proposition B.4.4 is satisfied. We proceed to check Assumption (2), which concerns the approximation of F_N by $\mathsf{F}_{N,M}$ on average.

We write $\mathsf{E}_{N,M} = \mathsf{F}_N - \mathsf{F}_{N,M}$. The expectation of $|\mathsf{E}_{N,M}|$ is given by

$$\begin{aligned}\mathbf{E}_N(|\mathsf{E}_{N,M}|) &= \frac{1}{N}\sum_{n\leqslant N}\left|\prod_{p|n}\left(1-\frac{1}{p}\right) - \prod_{\substack{p|n\\ p\leqslant M}}\left(1-\frac{1}{p}\right)\right| \\ &\leqslant \frac{1}{N}\sum_{n\leqslant N}\left|\prod_{\substack{p|n\\ p>M}}\left(1-\frac{1}{p}\right) - 1\right|.\end{aligned}$$

For a given n, expanding the product, we see that the quantity

$$\prod_{\substack{p|n\\ p>M}}\left(1-\frac{1}{p}\right) - 1$$

is bounded by the sum of $1/d$ over integers $d \geqslant 2$ which are squarefree, divide n, and have all prime factors $> M$; let D_n be the set of such integers. In particular, we always have $M < d \leqslant N$ if $d \in D_n$.

[4] Note that a similar argument reappears in a much more sophisticated context in Chapter 5 (see the proof of Theorem 5.2.2).

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

Thus

$$\mathbf{E}_{\mathrm{N}}(|\mathsf{E}_{\mathrm{N},\mathrm{M}}|) \leqslant \frac{1}{\mathrm{N}} \sum_{n \leqslant \mathrm{N}} \sum_{d \in \mathrm{D}_n} \frac{1}{d} \leqslant \sum_{\mathrm{M} < d \leqslant \mathrm{N}} \frac{1}{d} \times \frac{1}{\mathrm{N}} \sum_{\substack{n \leqslant \mathrm{N} \\ n \equiv 0 \,(\mathrm{mod}\, d)}} 1$$

$$\leqslant \sum_{\mathrm{M} < d \leqslant \mathrm{N}} \frac{1}{d^2} \leqslant \frac{1}{\mathrm{M}}$$

for all $\mathrm{N} \geqslant \mathrm{M}$. Assumption (2) of Proposition B.4.4 follows immediately, and we conclude that $(\mathsf{F}_{\mathrm{N}})_{\mathrm{N} \geqslant 1}$ converges in law, and that its limit is the limit in law F of the random variables F_{M} as $\mathrm{M} \to +\infty$. The last thing to check in order to finish the proof is that the random product

$$\prod_p \left(1 - \frac{\mathrm{B}_p}{p}\right) \tag{1.8}$$

over primes converges almost surely, and has the same law as F. The almost sure convergence follows from Kolmogorov's Three Series Theorem, applied to the logarithm of this product, which is a sum

$$\sum_p \mathrm{Y}_p, \qquad \mathrm{Y}_p = \log\left(1 - \frac{\mathrm{B}_p}{p}\right)$$

of independent random variables. Note that $\mathrm{Y}_p \leqslant 0$ and that it only takes the values 0 (with probability $1 - 1/p$) and $\log(1 - 1/p)$ (with probability $1/p$), so that

$$\mathbf{E}(\mathrm{Y}_p) = \frac{1}{p} \log\left(1 - \frac{1}{p}\right) \sim -\frac{1}{p^2},$$

$$\mathbf{V}(\mathrm{Y}_p) = \mathbf{E}(\mathrm{Y}_p^2) - \mathbf{E}(\mathrm{Y}_p)^2 = \frac{1}{p} \log\left(1 - \frac{1}{p}\right)^2 - \frac{1}{p^2} \log\left(1 - \frac{1}{p}\right)^2 \ll \frac{1}{p^3},$$

which implies by Theorem B.10.1 that the random series $\sum \mathrm{Y}_p$ converges almost surely, and hence so does its exponential, which is the product (1.8). Now, from this convergence almost surely, it is immediate that the law of the random product is also the law of F. □

In Section 2.2 of the next chapter, we will state and prove a theorem due to Erdős and Wintner that implies the existence of limiting distributions for much more general multiplicative functions.

Remark 1.4.2 The distribution function of the arithmetic function $n \mapsto \varphi(n)/n$ is the function defined for $x \in \mathbf{R}$ by

$$f(x) = \mathbf{P}(\mathrm{F} \leqslant x).$$

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

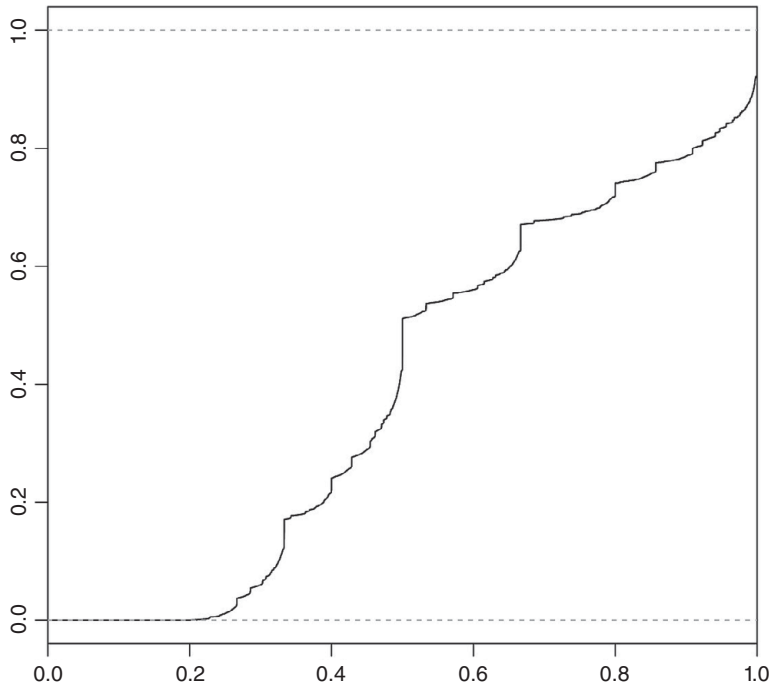

Figure 1.1 Empirical plot of the distribution function of $\varphi(n)/n$ for $n \leqslant 10^6$.

This function has been extensively studied, and is still the object of current research. It is a concrete example of a function exhibiting unusual properties in real analysis: it was proved by Schoenberg [108, 109] that f is continuous and strictly increasing, and by Erdős [34] that it is purely singular, that is, that there exists a set N of Lebesgue measure 0 in $\mathbf{R}$ such that $\mathbf{P}(\mathsf{F} \in \mathrm{N}) = 1$; this means that the function f is differentiable for all $x \notin \mathrm{N}$, with derivative equal to 0 (Exercise 1.4.4 explains the proof).

In Figure 1.1, we plot the "empirical" values of f coming from integers $n \leqslant 10^6$.

In the next two exercises, we use the notation of Proposition 1.4.1.

Exercise 1.4.3 Prove probabilistically that

$$\lim_{\mathrm{N}\to+\infty} \mathbf{E}_{\mathrm{N}}(\mathsf{F}_{\mathrm{N}}) = \frac{1}{\zeta(2)} \quad \text{and}$$
$$\lim_{\mathrm{N}\to+\infty} \mathbf{E}_{\mathrm{N}}(\mathsf{F}_{\mathrm{N}}^{-1}) = \prod_p \left(1 + \frac{1}{p(p-1)}\right) = \frac{\zeta(2)\zeta(3)}{\zeta(6)},$$

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

where

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}$$

is the Riemann zeta function (see Corollary C.1.5 for the product expression). In other words, we have

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n\leqslant N} \frac{\varphi(n)}{n} = \frac{1}{\zeta(2)} \quad \text{and} \quad \lim_{N\to+\infty} \frac{1}{N} \sum_{n\leqslant N} \frac{n}{\varphi(n)} = \frac{\zeta(2)\zeta(3)}{\zeta(6)}.$$

Recover these formulas using Möbius inversion (as in the "direct" proof of Proposition 1.3.3).

Exercise 1.4.4 (1) Prove that the support of the law of F is $[0,1]$. [**Hint**: Use Proposition B.10.8.]

By the Jessen–Wintner Purity Theorem (see, e.g., [20, Th. 3.26]), this fact implies that the function f is purely singular (in the sense of Remark 1.4.2), provided there exists a set N of Lebesgue measure 0 such that $\mathbf{P}(F \in N) > 0$. In turn, by elementary properties of absolutely continuous probability measures, this follows if there exists $\alpha > 0$ and, for any $\varepsilon > 0$, a Borel set $I_\varepsilon \subset [0,1]$ such that

(1) we have $\mathbf{P}(F \in I_\varepsilon) \geqslant \alpha$ for all ε small enough; and
(2) the Lebesgue measure of I_ε tends to 0 as $\varepsilon \to 0$.

The next questions will establish the existence of such sets. We define $G = \log(F)$, and for $M \geqslant 2$, we let G_M denote the partial sum

$$G_M = \sum_{p\leqslant M} \log\left(1 - \frac{B_p}{p}\right).$$

(2) Prove that for any $\delta > 0$, we have

$$\mathbf{P}(|G - G_M| > \delta) \ll \frac{1}{\delta M}$$

for any $M > 0$.

(3) For any finite set T of primes $p \leqslant M$, with characteristic function χ_T, prove that

$$\mathbf{P}(B_p = \chi_T(p) \text{ for } p \leqslant M) \gg \frac{1}{\log M} \times \prod_{p\in T} \frac{1}{p}.$$

Hint: Use the Mertens Formula (Proposition C.3.1).

(4) Let $\mathscr{T}_M$ be a set of subsets T of the set of primes $p \leqslant M$, and let X_M be the event

$$\{ \text{there exists } T \in \mathscr{T}_M \text{ such that } B_p = \chi_T(p) \text{ for } p \leqslant M \}.$$

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

Show that

$$\mathbf{P}(\mathrm{X_M}) \gg \frac{1}{\log \mathrm{M}} \sum_{\mathrm{T} \in \mathscr{T}_\mathrm{M}} \prod_{p \in \mathrm{T}} \frac{1}{p}.$$

(5) Let $\delta > 0$ be some auxiliary parameter and

$$\mathrm{I_M} = \bigcup_{\mathrm{T} \in \mathscr{T}_\mathrm{M}} \left[\sum_{p \in \mathrm{T}} \log(1 - 1/p) - \delta, \sum_{p \in \mathrm{T}} \log(1 - 1/p) + \delta, \right].$$

Show that the Lebesgue measure of $\mathrm{I_M}$ is $\leqslant 2\delta |\mathscr{T}_\mathrm{M}|$ and that

$$\mathbf{P}(\mathrm{G} \in \mathrm{I_M}) \gg \frac{1}{\log \mathrm{M}} \sum_{\mathrm{T} \in \mathscr{T}_\mathrm{M}} \prod_{p \in \mathrm{T}} \frac{1}{p} - \frac{1}{\delta \mathrm{M}}.$$

(6) Conclude by finding a choice of $\delta > 0$ and $\mathscr{T}_\mathrm{M}$ such that the Lebesgue measure of $\mathrm{I_M}$ tends to 0 as $\mathrm{M} \to +\infty$ whereas $\mathbf{P}(\mathrm{G} \in \mathrm{I_M}) \gg 1$ for M large enough.

1.5 Generalizations

Theorem 1.3.1 and Proposition 1.3.7 are obviously very simple statements. However, Proposition 1.4.1 has already shown that they should not be disregarded as trivial (and our careful presentation should – maybe – not be considered as overly pedantic). A further and even stronger sign in this direction is the fact that if one considers other natural sequences of probability measures on the integers, instead of the uniform measures on $\{1, \dots, \mathrm{N}\}$, one quickly encounters very delicate questions, and indeed fundamental open problems.

We have already mentioned the generalization related to polynomial values $\mathrm{P}(n)$ for some fixed polynomial $\mathrm{P} \in \mathbf{Z}[\mathrm{X}]$. Here are two other natural sequences of measures that have been studied.

1.5.1 Primes

Maybe the most important variant consists in replacing the space Ω_N of positive $n \leqslant \mathrm{N}$ by the subset Π_N of prime numbers $p \leqslant \mathrm{N}$ (with the uniform probability measure on these finite sets). According to the Prime Number Theorem (Theorem C.3.3), there are about $\mathrm{N}/(\log \mathrm{N})$ primes in Π_N. In this case, the qualitative analogue of Theorem 1.3.1 is given by the theorem of Dirichlet, Hadamard and de la Vallée Poussin on primes in arithmetic

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

progression (Theorem C.3.7), which implies that, for any fixed $q \geqslant 1$, the random variables π_q on Π_N converge in law to the probability measure on $\mathbf{Z}/q\mathbf{Z}$ which is the uniform measure on the subset $(\mathbf{Z}/q\mathbf{Z})^\times$ of invertible residue classes (this change of the measure compared with the case of integers is simply due to the obvious fact that at most one prime may be divisible by the integer q).

It is *expected* that a bound similar to (1.1) should be true. More precisely, there *should* exist a constant $C \geqslant 0$ such that

$$\left|\mathbf{E}_{\Pi_N}(f(\pi_q)) - \mathbf{E}(f)\right| \leqslant \frac{C(\log qN)^2}{\sqrt{N}}\|f\|_1, \tag{1.9}$$

but that statement, once it is translated to more standard notation, is very close to the Generalized Riemann Hypothesis for Dirichlet L-functions (which is Conjecture C.5.8).[5] Even a similar bound with $\sqrt{N}$ replaced by N^θ for any fixed $\theta > 0$ is not known, and would be a sensational breakthrough. Note that here the function f is defined on $(\mathbf{Z}/q\mathbf{Z})^\times$ and we have

$$\mathbf{E}(f) = \frac{1}{\varphi(q)} \sum_{a \in (\mathbf{Z}/q\mathbf{Z})^\times} f(a),$$

with $\varphi(q) = |(\mathbf{Z}/q\mathbf{Z})^\times|$ denoting the Euler function (see Example C.1.8).

However, weaker versions of (1.9), amounting roughly to a version valid on average over $q \leqslant \sqrt{N}$, are known: the Bombieri–Vinogradov Theorem states that, for any constant $A > 0$, there exists $B > 0$ such that we have

$$\sum_{q \leqslant \sqrt{N}/(\log N)^B} \max_{a \in (\mathbf{Z}/q\mathbf{Z})^\times} \left|\mathbf{P}_{\Pi_N}(\pi_q = a) - \frac{1}{\varphi(q)}\right| \ll \frac{1}{(\log N)^A}, \tag{1.10}$$

where the implied constant depends only on A (see, e.g., [59, Ch. 17]). In many applications, this is essentially as useful as (1.9).

Exercise 1.5.1 Compute the "probability" that $p - 1$ be squarefree, for p prime. (This can be done using the Bombieri–Vinogradov Theorem, for instance.)

[Further references: Friedlander and Iwaniec [43]; Iwaniec and Kowalski [59].**]**

1.5.2 Random walks

A more recent (and extremely interesting) type of problem arises from taking measures on $\mathbf{Z}$ derived from *random walks* on certain discrete groups.

[5] It implies it for nontrivial Dirichlet characters.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

For simplicity, we only consider a special case. Let $m \geqslant 2$ be an integer, and let $\mathrm{G} = \mathrm{SL}_m(\mathbf{Z})$ be the group of $m \times m$ matrices with integral coefficients and determinant 1. This is a complicated infinite (countable) group, but it is known to have finite generating sets. We fix one such set S, and assume that $1 \in \mathrm{S}$ and $\mathrm{S} = \mathrm{S}^{-1}$ for convenience. (A well-known example is the set S consisting of 1 and the elementary matrices $1 + \mathrm{E}_{i,j}$ for $1 \leqslant i \neq j \leqslant m$, where $\mathrm{E}_{i,j}$ is the matrix where only the (i,j)th coefficient is nonzero, and equal to 1, and their inverses $1 - \mathrm{E}_{i,j}$; the fact that these generate $\mathrm{SL}_n(\mathbf{Z})$ can be seen from the row-and-column operation reduction algorithm for such matrices.)

The generating set S defines then a random walk $(\gamma_n)_{n\geqslant 0}$ on G: let $(\xi_n)_{n\geqslant 1}$ be a sequence of independent S-valued random variables (defined on some probability space Ω) such that $\mathbf{P}(\xi_n = s) = 1/|\mathrm{S}|$ for all n and all $s \in \mathrm{S}$. Then we let

$$\gamma_0 = 1, \qquad \gamma_{n+1} = \gamma_n \xi_{n+1}.$$

Fix some (nonconstant) polynomial function F of the coefficients of an element $g \in \mathrm{G}$, so $\mathrm{F} \in \mathbf{Z}[(g_{i,j})]$ (for instance, $\mathrm{F}(g) = g_{1,1}$, or $\mathrm{F}(g) = \mathrm{Tr}(g)$ for $g = (g_{i,j})$ in G). We can then study the analogue of Theorem 1.3.1 when applied to the random variables $\pi_q(\mathrm{F}(\gamma_n))$ as $n \to +\infty$, or in other words, the distribution of $\mathrm{F}(g)$ modulo q, as g varies in G according to the distribution of the random walk.

Let $\mathrm{G}_q = \mathrm{SL}_m(\mathbf{Z}/q\mathbf{Z})$ be the finite special linear group. It is an elementary exercise, using finite Markov chains and the surjectivity of the projection map $\mathrm{G} \longrightarrow \mathrm{G}_q$, to check that the sequence of random variables $(\pi_q(\mathrm{F}(\gamma_n)))_{n\geqslant 0}$ converges in law as $n \to +\infty$. Indeed, its limit is a random variable F_q on $\mathbf{Z}/q\mathbf{Z}$ defined by

$$\mathbf{P}(\mathrm{F}_q = x) = \frac{1}{|\mathrm{G}_q|}|\{g \in \mathrm{G}_q \mid \mathrm{F}(g) = x\}|,$$

for all $x \in \mathbf{Z}/q\mathbf{Z}$, where we view F as also defining a function $\mathrm{F}\colon \mathrm{G}_q \longrightarrow \mathbf{Z}/q\mathbf{Z}$. (In other words, F_q is distributed like the direct image under F of the uniform measure on G_q.)

In fact, elementary Markov chain theory (or direct computations) shows that there exists a constant $c_q > 1$ such that for any function $f\colon \mathrm{G}_q \longrightarrow \mathbf{C}$, we have

$$\left|\mathbf{E}(f(\pi_q(\gamma_n)) - \mathbf{E}(f)\right| \leqslant \frac{\|f\|_1}{c_q^n}, \tag{1.11}$$

in analogy with (1.1), with

$$\|f\|_1 = \sum_{g\in \mathrm{G}_q} |f(g)|.$$

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

This is a very good result for a fixed q (note that the number of elements reached by the random walk after n steps also grows exponentially with n). For applications, our previous discussion already shows that it will be important to exploit (1.11) for q varying with n, and uniformly over a wide range of q. This requires an understanding of the variation of the constant c_q with q. It is a rather deep fact (Property (τ) of Lubotzky for $\mathrm{SL}_2(\mathbf{Z})$, and Property (T) of Kazhdan for $\mathrm{SL}_m(\mathbf{Z})$ if $m \geqslant 3$) that there exists $c > 1$, depending only on m, such that $c_q \geqslant c$ for all $q \geqslant 1$. Thus we do get a uniform bound

$$\left|\mathbf{E}(f(\pi_q(\gamma_n)) - \mathbf{E}(f)\right| \leqslant \frac{\|f\|_1}{c^n}$$

valid for all $n \geqslant 1$ and all $q \geqslant 1$. This is related to the theory (and applications) of *expander graphs*.

[Further references: Breuillard and Oh [21], Kowalski [65], [67].**]**

1.6 Outline of the Book

Here is now a quick outline of the main results that we will prove in the text. For detailed statements, we refer to the introductory sections of the corresponding chapters.

Chapter 2 presents first the Erdős–Wintner Theorem on the limiting distribution of additive functions, before discussing the Erdős–Kac Theorem. These are good examples to begin with, because they are the most natural starting point for probabilistic number theory, and remain quite lively topics of contemporary research. This will lead to natural appearances of the Gaussian distribution as well as Poisson distributions.

Chapters 3 and 4 are concerned with the distribution of values of the Riemann zeta function. We discuss results outside of the critical line (due to Bohr–Jessen, Bagchi and Voronin) in the first of these chapters, and consider deeper results on the critical line (due to Selberg, but following a recent presentation of Radziwiłł and Soundararajan) in the second. The limit theorems one obtains can have rather unorthodox limiting distributions (random Euler products, sometimes viewed as random functions, and – conjecturally – also eigenvalues of random unitary matrices of large size).

Chapter 5 takes up a fascinating topic in the distribution of prime numbers: the *Chebychev bias*, which attempts to compare the number of primes $\leqslant x$ in various residue classes modulo a fixed integer $q \geqslant 1$, and to see if some classes are "more equal" than others. Our treatment follows the basic paper of Rubinstein and Sarnak.

https://doi.org/10.1017/9781108888226.002 Published online by Cambridge University Press

In Chapter 6, we consider the distribution, in the complex plane, of polygonal paths joining partial sums of Kloosterman sums, following work of the author and W. Sawin [79, 12]. Here we will use convergence in law in Banach spaces and some elementary probability in Banach spaces, and the limit object that arises will be a very special random Fourier series.

In all of these chapters, we usually only discuss in detail one specific example of fairly general results and theories: just the additive function $\omega(n)$ instead of more general additive functions, just the Riemann zeta function instead of more general L-functions, and specific families of exponential sums. However, we will briefly mention some of the natural generalizations of the results presented.

Similarly, since our objective in this book is explicitly to write an *introduction* to the topic of probabilistic number theory, we did not attempt to cover the most refined results or the cutting-edge of research, or to discuss all possible topics. For the same reason, we do not discuss in depth the applications of our main results, although we usually mention at least some of them. Besides the discussion in Chapter 7 of other areas of interaction between probability theory and number theory, the reader is invited to read the short survey by Perret-Gentil [92].

At the end of the book are appendices that discuss the results of complex analysis, probability theory and number theory that we use in the main chapters of the book. In general, these are presented with some examples and detailed references, but without complete proofs, at least when they can be considered to be standard parts of their respective fields. We do not expect every reader to already be familiar with all of these facts, and in order to make it possible to read the text relatively linearly, each chapter begins with a list of the main results from these appendices that it will require, with the corresponding reference (when no reference is given, this means that the result in question will be presented within the chapter itself). We also note that the number-theoretic results in Appendix C are stated in the "classical" style of analytic number theory, without attempting to fit them to a probabilistic interpretation.