2

Classical Probabilistic Number Theory

Probability tools	Arithmetic tools
Definition of convergence in law (Section B.3)	Integers in arithmetic progressions (Section 1.3)
Convergence in law using auxiliary parameters (Prop. B.4.4)	Mertens and Chebychev estimate (Prop. C.3.1)
Central Limit Theorem (Th. B.7.2)	Additive and multiplicative functions (Section C.1, C.2)
Gaussian random variables (Section B.7)	
The method of moments (Th. B.5.5)	
Poisson random variables (Section B.9)	

2.1 Introduction

This chapter contains some of the earliest theorems of probabilistic number theory. We will prove the Erdős–Kac Theorem, but first we consider an even more classical topic: the distribution of multiplicative and additive arithmetic functions. The essential statements predate the Erdős–Kac Theorem, and can be taken to be the beginning of true probabilistic number theory. As we will see, the limiting distributions that are obtained are far from generic.

2.2 Distribution of Arithmetic Functions

The classical problem of the distribution of the values of arithmetic functions concerns the limiting behavior of (arithmetic) random variables of the

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

form $g(\mathsf{S}_N)$, where g is an additive or multiplicative function, and S_N is the identity random variable on the probability space $\Omega_N = \{1, \dots, N\}$ with uniform probability measure. We saw an example in Proposition 1.4.1, but we will now prove a much more general statement.

In fact, in the additive case (see Section C.2 for the definition of additive functions), there is a remarkable *characterization* of those additive functions g for which the sequence $(g(\mathsf{S}_N))_N$ converges in law as $N \to +\infty$. Arithmetically, it may be surprising that it depends on no more than Theorem 1.3.1 (or Corollary 1.3.9), and the simplest upper bound of the right order of magnitude for the numbers of primes less than a given quantity (Chebychev's estimate); this was not even needed for Proposition 1.4.1.

Theorem 2.2.1 *Let g be a complex-valued additive function such that the series*

$$\sum_{|g(p)|\leqslant 1} \frac{g(p)}{p}, \qquad \sum_{|g(p)|\leqslant 1} \frac{|g(p)|^2}{p}, \qquad \sum_{|g(p)|>1} \frac{1}{p}$$

converge. Then the sequence of random variables $(g(\mathsf{S}_N))_N$ converges in law to the series over primes

$$\sum_p g(p^{\mathrm{V}_p}), \tag{2.1}$$

where $(\mathrm{V}_p)_p$ is a sequence of independent geometric random variables with

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

for $k \geqslant 0$.

Recall that, in terms of p-adic valuations of integers, we can write

$$g(n) = \sum_p g\left(p^{v_p(n)}\right)$$

for any integer $n \geqslant 1$. Since the sequence of p-adic valuations converges in law to the sequence (V_p) (Corollary 1.3.9), the formula (2.1) for the limiting distribution appears as a completely natural expression.

Proof We write $g = g^\flat + g^\sharp$ where both summands are additive functions, and

$$g^\flat(p^k) = \begin{cases} g(p) & \text{if } k = 1 \text{ and } |g(p)| \leqslant 1, \\ 0 & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

Thus $g^\sharp(p) = 0$ for a prime p unless $|g(p)| > 1$. We denote by (B_p) the Bernoulli random variable indicator function of the event $\{\mathrm{V}_p = 1\}$; we have

$$\mathbf{P}(\mathrm{B}_p = 1) = \frac{1}{p}\left(1 - \frac{1}{p}\right).$$

We will prove that the vectors $(g^\flat(\mathsf{S}_\mathrm{N}), g^\sharp(\mathsf{S}_\mathrm{N}))$ converge in law to

$$\left(\sum_p g^\flat(p^{\mathrm{V}_p}), \sum_p g^\sharp(p^{\mathrm{V}_p})\right),$$

and the desired conclusion then follows by composing with the continuous addition map $\mathbf{C}^2 \to \mathbf{C}$ (i.e., applying Proposition B.3.2).

We will apply Proposition B.4.4 to the random vectors $\mathsf{G}_\mathrm{N} = (g^\flat(\mathsf{S}_\mathrm{N}), g^\sharp(\mathsf{S}_\mathrm{N}))$ (with values in $\mathbf{C}^2$), with the approximations $\mathsf{G}_\mathrm{N} = \mathsf{G}_{\mathrm{N},\mathrm{M}} + \mathsf{E}_{\mathrm{N},\mathrm{M}}$, where

$$\mathsf{G}_{\mathrm{N},\mathrm{M}} = \left(\sum_{p\leqslant \mathrm{M}} g^\flat(p^{v_p(\mathsf{S}_\mathrm{N})}), \sum_{p\leqslant \mathrm{M}} g^\sharp(p^{v_p(\mathsf{S}_\mathrm{N})})\right).$$

Let $\mathrm{M} \geqslant 1$ be fixed. The random vectors $\mathsf{G}_{\mathrm{N},\mathrm{M}}$ are finite sums, and are expressed as obviously continuous functions of the valuations v_p of the elements of Ω_N, for $p \leqslant \mathrm{M}$. Since the vector of these valuations converges in law to $(\mathrm{V}_p)_{p\leqslant \mathrm{M}}$ by Corollary 1.3.9, applying composition with a continuous map (Proposition B.3.2 again), it follows that $(\mathsf{G}_{\mathrm{N},\mathrm{M}})_\mathrm{N}$ converges in law as $\mathrm{N} \to +\infty$ to the vector

$$\left(\sum_{p\leqslant \mathrm{M}} g^\flat(p^{\mathrm{V}_p}), \sum_{p\leqslant \mathrm{M}} g^\sharp(p^{\mathrm{V}_p})\right).$$

It is therefore enough to verify that Assumption (2) of Proposition B.4.4 holds, and we may do this separately for each of the two coordinates of the vector (by taking the norm on $\mathbf{C}^2$ in the proposition to be the maximal of the modulus of the two coordinates).

We begin with the second coordinate involving $g^\sharp$. For any $\delta > 0$, and $2 \leqslant \mathrm{M} < \mathrm{N}$, we have

$$\begin{aligned}\mathbf{P}_\mathrm{N}\left(\left|\sum_{\mathrm{M}<p\leqslant \mathrm{N}} g^\sharp(p^{v_p(\mathsf{S}_\mathrm{N})})\right| > \delta\right) &\leqslant \sum_{\mathrm{M}<p\leqslant \mathrm{N}} \mathbf{P}_\mathrm{N}(v_p(\mathsf{S}_\mathrm{N}) \geqslant 2) \\ &\quad + \sum_{\substack{\mathrm{M}<p\leqslant \mathrm{N}\\ |g(p)|>1}} \mathbf{P}_\mathrm{N}(v_p(\mathsf{S}_\mathrm{N}) = 1) \\ &\leqslant \sum_{p>\mathrm{M}} \frac{1}{p^2} + \sum_{\substack{p>\mathrm{M}\\ |g(p)|>1}} \frac{1}{p} \qquad (2.2)\end{aligned}$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

(simply because, if the sum is nonzero, at least one term must be nonzero, and the probability of a union of countably many sets is bounded by the sums of the probabilities of the individual sets).

Since the right-hand side converges to 0 as $M \to +\infty$ (by assumption), this verifies that the variant discussed in Remark B.4.5 of the assumption of Proposition B.4.4 holds (note that the series

$$\sum_{p \leqslant M} g^\sharp(p^{V_p})$$

converges in law by a straightforward application of Kolmogorov's Three Series Theorem, which is stated in Remark B.10.2 – indeed, since $|g^\sharp| \geqslant 1$, it suffices to observe that

$$\sum_{p \leqslant M} \mathbf{P}(|g^\sharp(p_p^V)| \geqslant 2) < +\infty,$$

which follows by arguing as in (2.2)).

We next handle $g^\flat$. We denote by $\mathsf{B}_{N,p}$ the Bernoulli random variable indicator of the event $\{v_p(\mathsf{S}_N) = 1\}$, and define

$$\varpi_N(p) = \mathbf{P}_N(\mathsf{B}_{N,p} = 1) = \mathbf{P}_N(v_p(\mathsf{S}_N) = 1).$$

We also write $\varpi(p) = \mathbf{P}(\mathrm{B}_p = 1)$. Note that

$$\varpi_N(p) \leqslant \frac{1}{p} \quad \text{and} \quad \varpi_N(p) = \frac{1}{p}\left(1 - \frac{1}{p}\right) + \mathrm{O}\left(\frac{1}{N}\right) = \varpi(p) + \mathrm{O}\left(\frac{1}{N}\right).$$

The first coordinate of $\mathsf{E}_{N,M}$ is

$$\mathsf{H}_{N,M} = \sum_{p>M} g^\flat(p^{V_p}) = \sum_{p>M} g^\flat(p)\mathsf{B}_{N,p}$$

(which is a finite sum, so convergence issues do not arise). We will prove that

$$\lim_{M\to+\infty} \limsup_{N\to+\infty} \mathbf{E}_N(|\mathsf{H}_{N,M}|^2) = 0,$$

which will also us to conclude.

By expanding the square, we have

$$\begin{aligned}\mathbf{E}_N(|\mathsf{H}_{N,M}|^2) &= \mathbf{E}_N\left(\left|\sum_{p>M} g^\flat(p)\mathsf{B}_{N,p}\right|^2\right)\\ &= \sum_{p_1,p_2>M} \mathbf{E}_N\left(\overline{g^\flat(p_1)}g^\flat(p_2)\mathsf{B}_{N,p_1}\mathsf{B}_{N,p_2}\right). \qquad (2.3)\end{aligned}$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

The contribution of the diagonal terms $p_1 = p_2$ to (2.3) is

$$\sum_{p>\mathrm{M}} |g^\flat(p)|^2 \varpi_{\mathrm{N}}(p) \leqslant \sum_{p>\mathrm{M}} \frac{|g^\flat(p)|^2}{p}.$$

We have

$$\mathbf{E}_{\mathrm{N}}(\mathsf{B}_{\mathrm{N},p_1}\mathsf{B}_{\mathrm{N},p_2}) = \mathbf{P}_{\mathrm{N}}(v_{p_1}(\mathsf{S}_{\mathrm{N}}) = v_{p_2}(\mathsf{S}_{\mathrm{N}}) = 1) = \varpi(p_1)\varpi(p_2) + \mathrm{O}\left(\frac{1}{\mathrm{N}}\right)$$

(by Example 1.3.10), so that the nondiagonal terms become

$$\sum_{\substack{p_1,p_2>\mathrm{M}\\ p_1\neq p_2}} \overline{g^\flat(p_1)} g^\flat(p_2)\varpi(p_1)\varpi(p_2) + \mathrm{O}\Bigg(\frac{1}{\mathrm{N}} \sum_{\substack{p_1,p_2>\mathrm{M}\\ p_1p_2\leqslant \mathrm{N}}} |g^\flat(p_1)||g^\flat(p_2)|\Bigg). \tag{2.4}$$

The first term S_1 in this sum is

$$\mathrm{S}_1 = \left|\sum_{p>\mathrm{M}} g^\flat(p)\varpi(p)\right|^2 - \sum_{p>\mathrm{M}} |g^\flat(p)|^2\varpi(p)^2 \leqslant \left|\sum_{p>\mathrm{M}} g^\flat(p)\varpi(p)\right|^2$$
$$= \left|\sum_{p>\mathrm{M}} \frac{g^\flat(p)}{p}\left(1-\frac{1}{p}\right)\right|^2,$$

where the right-hand side of the last equality is convergent because of the assumptions of the theorem, so that the left-hand side is also finite.

Next, since $|g^\flat(p)| \leqslant 1$ for all primes, the second term S_2 in (2.4) satisfies

$$\mathrm{S}_2 \ll \frac{1}{\mathrm{N}} \sum_{\substack{p_1,p_2>\mathrm{M}\\ p_1p_2\leqslant \mathrm{N}}} 1 \ll \frac{\log\log \mathrm{N}}{\log \mathrm{N}}$$

for all $\mathrm{M} \geqslant 1$ by Chebychev's estimate of Proposition C.3.1 (extended to products of two primes as in Exercise C.3.2 (2)). Finally, from the convergence assumptions, this means that

$$\limsup_{\mathrm{N}\to+\infty} \mathbf{E}_{\mathrm{N}}(|\mathsf{H}_{\mathrm{N},\mathrm{M}}|^2) \ll \left|\sum_{p>\mathrm{M}} \frac{g^\flat(p)}{p}\right|^2 + \sum_{p>\mathrm{M}} \frac{|g^\flat(p)|^2}{p} \to 0$$

as $\mathrm{M} \to +\infty$, and this concludes the proof. □

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

Remark 2.2.2 The result above is due to Erdős [33]; the fact that the converse assertion also holds (namely, that if the sequence $(g(\mathsf{S}_N))_N$ converges in law, then the three series

$$\sum_{|g(p)|\leqslant 1} \frac{g(p)}{p}, \qquad \sum_{|g(p)|\leqslant 1} \frac{|g(p)|^2}{p}, \qquad \sum_{|g(p)|>1} \frac{1}{p}$$

are convergent) is known as the Erdős–Wintner Theorem [36]. The reader may be interested in thinking about proving this; see, for example, [115, pp. 327–328] for the details.

Although it is of course customary and often efficient to pass from additive functions to multiplicative functions by taking the logarithm, this is not always possible. For instance, the (multiplicative) Möbius function $\mu(n)$ *does* have the property that the sequence $(\mu(\mathsf{S}_N))_N$ converges in law to a random variable taking values 0, 1 and -1 with probabilities which are equal, respectively, to

$$1 - \frac{6}{\pi^2}, \qquad \frac{3}{\pi^2}, \qquad \frac{3}{\pi^2}.$$

The limiting probability that $\mu(n) = 0$ comes from the elementary Proposition 1.3.3, but the fact that, among the values 1 and -1, the asymptotic probability is equal, is quite a bit deeper: it turns out to be "elementarily" equivalent to the Prime Number Theorem in the form

$$\pi(x) \sim \frac{x}{\log x}$$

as $x \to +\infty$ (see, e.g., [59, §2.1] for the proof). However, there is no additive function $\log \mu(n)$, so we cannot even begin to speak of its potential limiting distribution!

2.3 The Erdős–Kac Theorem

We begin by recalling the statement (see Theorem 1.1.1), in its probabilistic phrasing:

Theorem 2.3.1 (Erdős–Kac Theorem) *For* $N \geqslant 1$, *let* $\Omega_N = \{1, \dots, N\}$ *with the uniform probability measure* $\mathbf{P}_N$. *Let* X_N *be the random variable*

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

on Ω_N *for* $N \geqslant 3$. *Then* $(\mathsf{X}_N)_{N\geqslant 3}$ *converges in law to a standard Gaussian random variable, that is, to a Gaussian random variable with expectation* 0 *and variance* 1.

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

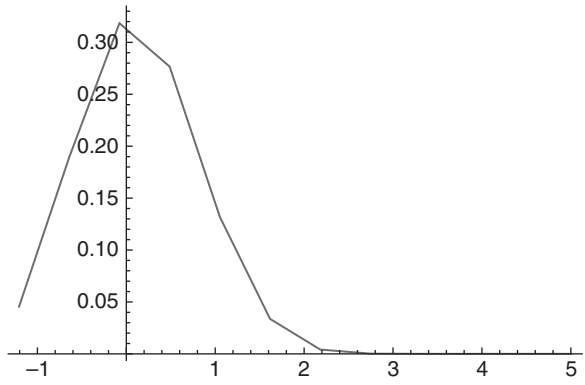

Figure 2.1 The normalized number of prime divisors for $n \leqslant 10^{10}$.

Figure 2.1 shows a plot of the empirical density of X_{N} for $\mathrm{N} = 10^{10}$: one can see something that could be the shape of the Gaussian density appearing, but the fit is very far from perfect (we will comment later why this could be expected).

The original proof of Theorem 2.3.1 is due to Erdős and Kac in 1939 [35]. We will explain a proof following the work of Granville and Soundararajan [51] and of Billingsley [9, p. 394]. As usual, the presentation emphasizes the probabilistic nature of the argument.

As before, we begin by explaining why the statement can be considered to be unsurprising. This is an elaboration of the type of heuristic argument that we used to justify the limit in Theorem 2.2.1.

The arithmetic function ω is additive. Write

$$\omega(n) = \sum_p \mathsf{B}_p(n)$$

for $n \in \Omega_{\mathrm{N}}$, where B_p is as usual the Bernoulli random variable on Ω_{N} that is the characteristic function of the event $p \mid n$. Using Proposition 1.3.7, the natural probabilistic guess for a limit (if there was one) would be the series

$$\sum_p \mathrm{B}_p,$$

where (B_p) are independent Bernoulli random variables, as in Proposition 1.4.1. But this series diverges almost surely: indeed, the series

$$\sum_p \mathbf{E}(\mathrm{B}_p) = \sum_p \frac{1}{p}$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

diverges by the basic Mertens estimate from prime number theory, namely,

$$\sum_{p \leqslant \mathrm{N}} \frac{1}{p} = \log\log \mathrm{N} + \mathrm{O}(1)$$

for $\mathrm{N} \geqslant 3$ (see Proposition C.3.1 in Appendix C), so that the divergence follows from Kolmogorov's Theorem, B.10.1 (or indeed an application of the Borel–Cantelli Lemma; see Exercise B.10.4).

One can however refine the formula for ω by observing that $n \in \Omega_{\mathrm{N}}$ has no prime divisor larger than N, so that we also have

$$\omega(n) = \sum_{p \leqslant \mathrm{N}} \mathsf{B}_p(n) \tag{2.5}$$

for $n \in \Omega_{\mathrm{N}}$. Correspondingly, we may expect that the probabilistic distribution of ω on Ω_{N} will be similar to that of the sum

$$\sum_{p \leqslant \mathrm{N}} \mathrm{B}_p. \tag{2.6}$$

But the latter is a sum of *independent* (though not identically distributed) random variables, and its asymptotic behavior is therefore well understood. In fact, a simple case of the Central Limit Theorem (see Theorem B.7.2) implies that the renormalized random variables

$$\frac{\displaystyle\sum_{p \leqslant \mathrm{N}} \mathrm{B}_p - \sum_{p \leqslant \mathrm{N}} p^{-1}}{\displaystyle\sqrt{\sum_{p \leqslant \mathrm{N}} p^{-1}(1 - p^{-1})}}$$

converge in law to a standard Gaussian random variable. It is then to be expected that the arithmetic sums (2.5) are sufficiently close to (2.6) so that a similar renormalization of ω on Ω_{N} will lead to the same limit, and this is exactly the statement of Theorem 2.3.1 (by the Mertens Formula again).

We now begin the rigorous proof. We will prove convergence in law using the method of moments, as explained in Section B.3 of Appendix B, specifically in Theorem B.5.5 and Remark B.5.9. This is definitely not the only way to confirm the heuristic above, but it may be the simplest.

More precisely, we will proceed as follows:

(1) We show, using Theorem 1.3.1, that for any fixed integer $k \geqslant 0$, we have

$$\mathbf{E}_{\mathrm{N}}(\mathsf{X}_{\mathrm{N}}^k) = \mathbf{E}(\mathrm{X}_{\mathrm{N}}^k) + o(1),$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

where (X_N) is the same renormalized random variable described above, namely,

$$X_N = \frac{Z_N - \mathbf{E}(Z_N)}{\sqrt{\mathbf{V}(Z_N)}}$$

with

$$Z_N = \sum_{p \leqslant N} B_p. \tag{2.7}$$

(2) As we already mentioned, the Central Limit Theorem applies to the sequence (X_N), and shows that it converges in law to a standard Gaussian random variable $\mathcal{N}$.
(3) It follows that

$$\lim_{N \to +\infty} \mathbf{E}_N(X_N^k) = \mathbf{E}(\mathcal{N}^k),$$

and hence, by the method of moments (Theorem B.5.5), we conclude that X_N converges in law to $\mathcal{N}$. (Interestingly, we do not need to know the value of the moments $\mathbf{E}(\mathcal{N}^k)$ for this argument to apply.)

This sketch indicates that the Erdős–Kac Theorem is really a result of very general nature that should be valid for many random integers, and not merely for a uniformly chosen integer in Ω_N. Note that only Step 1 has real arithmetic content. As we will see, that arithmetic content is concentrated on two results: Theorem 1.3.1, which makes the link with probability theory, and the Mertens estimate, which is only required in the form of the divergence of the series

$$\sum_p \frac{1}{p}$$

(at least if one is ready to use its partial sums

$$\sum_{p \leqslant N} \frac{1}{p}$$

for renormalization, instead of the asymptotic value $\log \log N$).

We now implement this strategy. As will be seen, some tweaks will be required. (The reader is invited to check that omitting those tweaks leads, at the very least, to a much more complicated-looking problem!)

Step 1 (Truncation). This is a classical technique that applies here, and is used to shorten and simplify the sum in (2.7), in order to control the error terms

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

in the next step. We consider the random variables B_p on Ω_N as above, that is, $\mathsf{B}_p(n) = 1$ if p divides n and $\mathsf{B}_p(n) = 0$ otherwise. Let

$$\sigma_N = \sum_{p \leqslant N} \frac{1}{p}.$$

We only need recall at this point that $\sigma_N \to +\infty$ as $N \to +\infty$. We then define

$$Q = N^{1/(\log\log N)^{1/3}} \tag{2.8}$$

and

$$\tilde{\omega}(n) = \sum_{\substack{p \mid n \\ p \leqslant Q}} 1 = \sum_{p \leqslant Q} \mathsf{B}_p(n) \quad \text{and} \quad \tilde{\omega}_0(n) = \sum_{p \leqslant Q} \left(\mathsf{B}_p(n) - \frac{1}{p} \right)$$

viewed as random variables on Ω_N. The point of this truncation is the following: first, for $n \in \Omega_N$, we have

$$\tilde{\omega}(n) \leqslant \omega(n) \leqslant \tilde{\omega}(n) + (\log\log N)^{1/3},$$

simply because if $\alpha > 0$ and if $p_1, \ldots, p_m$ are primes $\geqslant N^{\alpha}$ dividing $n \leqslant N$, then we get

$$N^{m\alpha} \leqslant p_1 \cdots p_m \leqslant N,$$

and hence $m \leqslant \alpha^{-1}$. Second, for any $N \geqslant 1$ and any $n \in \Omega_N$, we get by definition of σ_N the identity

$$\begin{aligned} \tilde{\omega}_0(n) &= \tilde{\omega}(n) - \sum_{p \leqslant Q} \frac{1}{p} \\ &= \omega(n) - \sigma_N + O((\log\log N)^{1/3}) \end{aligned} \tag{2.9}$$

because the Mertens formula

$$\sum_{p \leqslant x} \frac{1}{p} = \log\log x + O(1)$$

and the definition of σ_N show that

$$\sum_{p \leqslant Q} \frac{1}{p} = \sum_{p \leqslant N} \frac{1}{p} + O(\log\log\log N) = \sigma_N + O(\log\log\log N).$$

Now define

$$\tilde{X}_N(n) = \frac{\tilde{\omega}_0(n)}{\sqrt{\sigma_N}}$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

as random variables on Ω_N. We will prove that $\tilde{\mathsf{X}}_N$ converges in law to $\mathcal{N}$. The elementary Lemma B.5.3 of Appendix B (applied using (2.9)) then shows that the random variables

$$n \mapsto \frac{\omega(n) - \sigma_N}{\sqrt{\sigma_N}}$$

converge in law to $\mathcal{N}$. Finally, applying the same lemma one more time using the Mertens formula we obtain the Erdős–Kac Theorem.

It remains to prove the convergence of $\tilde{\mathsf{X}}_N$. We fix a nonnegative integer k, and our target is to prove the limit

$$\mathbf{E}_N(\tilde{\mathsf{X}}_N^k) \to \mathbf{E}(\mathcal{N}^k) \tag{2.10}$$

as $N \to +\infty$. Once this is proved for all k, then the method of moments shows that (X_N) converges in law to the standard normal random variable $\mathcal{N}$.

Remark 2.3.2 We might also have chosen to perform a truncation at $p \leqslant N^\alpha$ for some fixed $\alpha \in]0,1[$. However, in that case, we would need to adjust the value of α depending on k in order to obtain (2.10), and then passing from the truncated variables to the original ones would require some minor additional argument. Note that the function $(\log\log N)^{1/3}$ which is used to define the truncation could be replaced by any function going to infinity slower than $(\log\log N)^{1/2}$.

Step 2 (Moment computation). We now begin the proof of (2.10). We use the definition of $\tilde{\omega}_0(n)$ and expand the kth power in $\mathbf{E}_N(\tilde{\mathsf{X}}_N^k)$ to derive

$$\mathbf{E}_N(\tilde{\mathsf{X}}_N^k) = \frac{1}{\sigma_N^{k/2}} \sum_{p_1 \leqslant Q} \cdots \sum_{p_k \leqslant Q} \mathbf{E}_N\left(\left(\mathsf{B}_{p_1} - \frac{1}{p_1}\right) \cdots \left(\mathsf{B}_{p_k} - \frac{1}{p_k}\right)\right)$$

(where we omit for simplicity the subscripts N for the arithmetic random variables B_{p_i}). The crucial point is that the random variable

$$\left(\mathsf{B}_{p_1} - \frac{1}{p_1}\right) \cdots \left(\mathsf{B}_{p_k} - \frac{1}{p_k}\right) \tag{2.11}$$

can be expressed as $f(\pi_q)$ for some modulus $q \geqslant 1$ and some function $f : \mathbf{Z}/q\mathbf{Z} \longrightarrow \mathbf{C}$, so that the basic result of Theorem 1.3.1 may be applied to each summand.

To be precise, the value at $n \in \Omega_N$ of the random variable (2.11) only depends on the residue class x of n in $\mathbf{Z}/q\mathbf{Z}$, where q is the least common multiple of $p_1, \ldots, p_k$. In fact, this value is equal to $f(x)$, where

$$f(x) = \left(\delta_{p_1}(x) - \frac{1}{p_1}\right) \cdots \left(\delta_{p_k}(x) - \frac{1}{p_k}\right)$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

with δ_{p_i} denoting the characteristic function of the residues classes modulo q which are 0 modulo p_i. It is clear that $|f(x)| \leqslant 1$, as product of terms which are all $\leqslant 1$, and hence we have the bound

$$\|f\|_1 \leqslant q$$

(this is extremely imprecise, but suffices for now). From this we get

$$\left|\mathbf{E}_{\mathrm{N}}\left(\left(\mathsf{B}_{p_1} - \frac{1}{p_1}\right)\cdots\left(\mathsf{B}_{p_k} - \frac{1}{p_k}\right)\right) - \mathbf{E}(f)\right| \leqslant \frac{2q}{\mathrm{N}} \leqslant \frac{2\mathrm{Q}^k}{\mathrm{N}}$$

by Theorem 1.3.1.

But by the definition of f, we also see that

$$\mathbf{E}(f) = \mathbf{E}\left(\left(\mathrm{B}_{p_1} - \frac{1}{p_1}\right)\cdots\left(\mathrm{B}_{p_k} - \frac{1}{p_k}\right)\right),$$

where the random variables (B_p) form a sequence of *independent* Bernoulli random variables with $\mathbf{P}(\mathrm{B}_p = 1) = 1/p$ (the (B_p) for p dividing q are realized concretely as the characteristic functions δ_p on $\mathbf{Z}/q\mathbf{Z}$ with uniform probability measure).

Therefore we derive

$$\begin{aligned}\mathbf{E}_{\mathrm{N}}(\tilde{\mathrm{X}}_{\mathrm{N}}^k) &= \frac{1}{\sigma_{\mathrm{N}}^{k/2}} \sum_{p_1\leqslant \mathrm{Q}} \cdots \sum_{p_k\leqslant \mathrm{Q}} \left\{\mathbf{E}\left(\left(\mathrm{B}_{p_1} - \frac{1}{p_1}\right)\cdots\left(\mathrm{B}_{p_k} - \frac{1}{p_k}\right)\right)\right.\\ &\qquad \left. + \mathrm{O}(\mathrm{Q}^k\mathrm{N}^{-1})\right\}\\ &= \left(\frac{\tau_{\mathrm{N}}}{\sigma_{\mathrm{N}}}\right)^{k/2} \mathbf{E}(\mathrm{X}_{\mathrm{N}}^k) + \mathrm{O}(\mathrm{Q}^{2k}\mathrm{N}^{-1})\\ &= \left(\frac{\tau_{\mathrm{N}}}{\sigma_{\mathrm{N}}}\right)^{k/2} \mathbf{E}(\mathrm{X}_{\mathrm{N}}^k) + o(1)\end{aligned}$$

by our choice (2.8) of Q, where

$$\mathrm{X}_{\mathrm{N}} = \frac{1}{\sqrt{\tau_{\mathrm{N}}}} \sum_{p\leqslant \mathrm{Q}} \left(\mathrm{B}_p - \frac{1}{p}\right).$$

and

$$\tau_{\mathrm{N}} = \sum_{p\leqslant \mathrm{Q}} \frac{1}{p}\left(1 - \frac{1}{p}\right) = \sum_{p\leqslant \mathrm{Q}} \mathbf{V}(\mathrm{B}_p)$$

Step 3 (Conclusion). We now note that the version of the Central Limit Theorem which is recalled in Theorem B.7.2 applies to the random variables (B_p), and implies precisely that X_{N} converges in law to $\mathcal{N}$. But moreover, the

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

sequence (X_N) satisfies the uniform integrability assumption in the converse of the method of moments (see Example B.5.7, applied to the variables $B_p - 1/p$, which are independent and bounded by 1), and hence we have in particular

$$\mathbf{E}(X_N^k) \longrightarrow \mathbf{E}(\mathcal{N}^k).$$

Since $\tau_N \sim \sigma_N$ by the Mertens formula, we deduce that $\mathbf{E}_N(\tilde{X}_N^k)$ converges also to $\mathbf{E}(\mathcal{N}^k)$, which was our desired goal (2.10).

Exercise 2.3.3 One can avoid appealing to the converse of the method of moments by directly using the combinatorics involved in proofs of the Central Limit Theorem based on moments, which directly imply the convergence of moments for (X_N). Find such a proof in this special case. (See, for instance, [9, p. 391]; note that one must then know what are the moments of Gaussian random variables,; these are recalled in Proposition B.7.3.)

Exercise 2.3.4 Consider the probability spaces $\Omega_N^\flat$ consisting of integers $1 \leqslant n \leqslant N$ that are squarefree, with the uniform probability measure. Prove a version of the Erdős–Kac Theorem for the number of prime factors of an element of $\Omega_N^\flat$.

Exercise 2.3.5 For an integer $N \geqslant 1$, let $m(N)$ denote the set of integers that occur in the multiplication table for integers $1 \leqslant n \leqslant N$:

$$m(N) = \{k = ab \mid 1 \leqslant a \leqslant N, \quad 1 \leqslant b \leqslant N\} \subset \Omega_{N^2}.$$

Prove that $\mathbf{P}_{N^2}(m(N)) \to 0$, that is, that

$$\lim_{N \to +\infty} \frac{|m(N)|}{N^2} = 0.$$

This result is the basic statement concerning the "multiplication table" problem of Erdős; the precise asymptotic behavior of $|m(N)|$ has been determined by K. Ford [41] (improving results of Tenenbaum): we have

$$\frac{|m(N)|}{N^2} \asymp (\log N)^{-\alpha} (\log\log N)^{-3/2},$$

where

$$\alpha = 1 - \frac{1 + \log\log 2}{\log 2}.$$

See also the work of Koukoulopoulos [64] for generalizations.

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

Exercise 2.3.6 Let $\Omega(n)$ be the number of prime divisors of an integer $n \geqslant 1$, counted *with* multiplicity (so $\Omega(12) = 3$).[1] Prove that

$$\mathbf{P}_{\mathrm{N}}\left(\Omega(n) - \omega(n) \geqslant (\log\log \mathrm{N})^{1/4}\right) \leqslant (\log\log \mathrm{N})^{-1/4},$$

and deduce that the random variables

$$n \mapsto \frac{\Omega(n) - \log\log \mathrm{N}}{\sqrt{\log\log \mathrm{N}}}$$

also converge in law to $\mathcal{N}$.

Exercise 2.3.7 Try to prove the Erdős–Kac Theorem using the same "approximation" approach used in the proof of the Erdős–Wintner Theorem; what seems to go wrong (suggesting – if not proving – that one really should use different tools)?

2.4 Convergence without Renormalization

One important point that is made clear by the proof of the Erdős–Kac Theorem is that, although one might think that a statement about the behavior of the number of prime factors of integers tells us something about the distribution of primes (which are those integers n with $\omega(n) = 1$), the Erdős–Kac Theorem *provides no such information*. This can be seen mechanically from the proof, where the truncation step means in particular that primes are simply discarded unless they are smaller than the truncation level Q, or intuitively from the fact that the statement itself implies that "most" integers of size about N have $\log\log \mathrm{N}$ prime factors. For instance, as $\mathrm{N} \to +\infty$, we have

$$\mathbf{P}_{\mathrm{N}}\left(|\omega(n) - \log\log \mathrm{N}| > a\sqrt{\log\log \mathrm{N}}\right) \longrightarrow \mathbf{P}(|\mathcal{N}| > a)$$
$$\leqslant \sqrt{\frac{2}{\pi}} \int_a^{+\infty} e^{-x^2/2} dx \leqslant e^{-a^2/4}.$$

The problem lies in the normalization used to obtain a definite theorem of convergence in law: this "crushes" to some extent the more subtle aspects of the distribution of values of $\omega(n)$, especially with respect to extreme values. One can however still study this function probabilistically, but one must use less generic methods, to go beyond the "universal" behavior given by the Central Limit Theorem. There are at least two possible approaches in this direction, and we now briefly survey some of the results.

[1] We only use this function in this section and hope that confusion with Ω_{N} will be avoided.

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

Both methods have in common a switch in probabilistic focus: instead of looking for a Gaussian approximation of a normalized version of $\omega(n)$, one looks for a *Poisson approximation* of the *un-normalized* function.

Recall (see also Section B.9 in Appendix B) that a Poisson distribution with real parameter $\lambda \geqslant 0$ satisfies

$$\mathbf{P}(\lambda = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

for any integer $k \geqslant 0$. It turns out that an inductive computation using the Prime Number Theorem leads to the asymptotic formula

$$\begin{aligned}\frac{1}{\mathrm{N}}|\{n \leqslant \mathrm{N} \mid \omega(n) = k\}| &\sim \frac{1}{(k-1)!}\frac{(\log\log \mathrm{N})^{k-1}}{\log \mathrm{N}}\\ &= e^{-\log\log \mathrm{N}}\frac{(\log\log \mathrm{N})^{k-1}}{(k-1)!}\end{aligned}$$

for any fixed integer $k \geqslant 1$. This suggests that a better probabilistic approximation to the arithmetic function $\omega(n)$ on Ω_{N} is a Poisson distribution with parameter $\log\log \mathrm{N}$. The Erdős–Kac Theorem would then be, in essence, a consequence of the simple fact that a sequence (X_n) of Poisson random variables with parameters $\lambda_n \to +\infty$ has the property that

$$\frac{\mathrm{X}_n - \lambda_n}{\sqrt{\lambda_n}} \to \mathcal{N}, \tag{2.12}$$

as explained in Proposition B.9.1. Figure 2.2 shows the density of the values of $\omega(n)$ for $n \leqslant 10^{10}$ and the corresponding Poisson density. (The values of the probabilities for consecutive integers are joined by line segments for readability.)

Remark 2.4.1 The fact that the approximation error in such a statement is typically of size comparable to $\lambda_n^{-1/2}$ explains why one can expect that the convergence to a Gaussian in the Erdős–Kac Theorem should be extremely slow, since in that case the normalizing factor is of size $\log\log \mathrm{N}$, and goes to infinity very slowly.

To give a rigorous meaning to these ideas of Poisson approximation of $\omega(n)$, one must first give a precise definition, which can not be a straightforward convergence property, because the parameter of the Poisson approximation is not fixed.

Harper [53] (to the author's knowledge) was the first to implement explicitly such an idea. He derived an explicit upper bound for the *total variation*

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

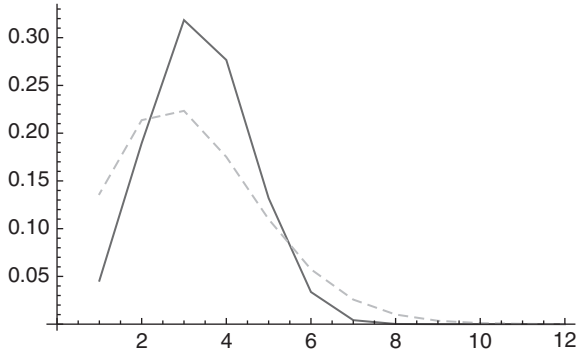

Figure 2.2 The number of prime divisors for $n \leqslant 10^{10}$ (solid line) compared with a Poisson distribution.

distance between a truncated version of $\omega(n)$ on Ω_N and a suitable Poisson random variable, namely, between

$$\sum_{\substack{p|n \\ p\leqslant Q}} 1, \quad \text{where } Q = N^{1/(3\log\log N)^2}$$

and a Poisson random variable Po_N with parameter

$$\lambda_N = \sum_{p\leqslant Q} \frac{1}{N}\left\lfloor \frac{N}{p} \right\rfloor$$

(so that the Mertens formula implies that $\lambda_N \sim \log\log N$).

Precisely, Harper proves that for *any* subset A of the nonnegative integers, we have

$$\left| \mathbf{P}_N\left(\sum_{\substack{p|n \\ p\leqslant Q}} 1 \in A \right) - \mathbf{P}(Po_N \in A) \right| \ll \frac{1}{\log\log N},$$

and moreover that the decay rate $(\log\log N)^{-1}$ is best possible. This requires some additional arithmetic information than the proof of Theorem 2.3.1 (essentially some form of sieve), but the arithmetic ingredients remain to a large extent elementary. On the other hand, new ingredients from probability theory are involved, especially cases of Stein's Method for Poisson approximation.

A second approach starts from a proof of the Erdős–Kac Theorem due to Rényi and Turán [100], which is the implementation of the Lévy Criterion for convergence in law. Precisely, they prove that

$$\mathbf{E}_N(e^{it\omega(n)}) = (\log N)^{e^{it}-1}(\Phi(t) + o(1)) \tag{2.13}$$

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

for any $t \in \mathbf{R}$ as $\mathrm{N} \to +\infty$ (in fact, uniformly for $t \in \mathbf{R}$ – note that the function here is 2π-periodic), with a factor $\Phi(t)$ given by

$$\Phi(t) = \frac{1}{\Gamma(e^{it})} \prod_p \left(1 - \frac{1}{p}\right)^{e^{it}} \left(1 + \frac{e^{it}}{p-1}\right), \tag{2.14}$$

where the product over all primes is absolutely convergent. Recognizing that the term $(\log \mathrm{N})^{e^{it}-1}$ is the characteristic function of a Poisson random variable Po_{N} with parameter $\log\log \mathrm{N}$, one can then obtain the Erdős–Kac Theorem by the same computation that leads to (2.12), combined with the continuity of Φ that shows that

$$\Phi\left(\frac{t}{\sqrt{\log\log \mathrm{N}}}\right) \longrightarrow \Phi(0) = 1$$

as $\mathrm{N} \to +\infty$.

The computation that leads to (2.13) is now interpreted as an instance of the Selberg–Delange method (see [115, II.5, Th. 3] for the general statement, and [115, II.6, Th. 1] for the special case of interest here).

It should be noted that the proof of (2.13) is quite a bit deeper than the proof of Theorem 2.3.1, and this is at it should be, because this formula contains precise information about the extreme values of $\omega(n)$, which we saw are not relevant to the Erdős–Kac Theorem. Indeed, taking $t = \pi$ and observing that $\Phi(\pi) = 0$ (because of the pole of the Gamma function), we obtain

$$\frac{1}{\mathrm{N}} \sum_{n \leqslant \mathrm{N}} (-1)^{\omega(n)} = \mathbf{E}(e^{-i\pi\omega(n)}) = o\left(\frac{1}{(\log \mathrm{N})^2}\right).$$

It is well known (as for the partial sums of the Möbius function, mentioned in Remark 2.2.2) that this implies elementarily the Prime Number Theorem

$$\sum_{p \leqslant \mathrm{N}} 1 \sim \frac{\mathrm{N}}{\log \mathrm{N}}$$

(see again [59, §2.1]).

The link between the formula (2.13) and Poisson distribution was noticed in joint work with Nikeghbali [77]. Among other things, we remarked that it implies easily a bound for the Kolmogorov–Smirnov distance between $n \mapsto \omega(n)$ on Ω_{N} and a Poisson random variable Po_{N}. Additional work with A. Barbour [5] leads to bounds in total variation distance, and to even better (but non-Poisson) approximations. Another suggestive remark is that if we consider

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

the independent random variables that appear in the proof of the Erdős–Kac theorem, namely,

$$\mathrm{X_N} = \sum_{p \leqslant \mathrm{N}} \left(\mathrm{B}_p - \frac{1}{p} \right),$$

where (B_p) is a sequence of independent Bernoulli random variables with $\mathbf{P}(\mathrm{B}_p = 1) = 1/p$, then we have (by a direct computation) the following analogue of (2.13):

$$\mathbf{E}(e^{it\mathrm{X_N}}) = (\log \mathrm{N})^{e^{it}-1} \left(\prod_p \left(1 - \frac{1}{p}\right)^{e^{it}} \left(1 + \frac{e^{it}}{p-1}\right) + o(1) \right).$$

It is natural to ask then if there is a similar meaning to the factor $1/\Gamma(e^{it})$ that also appears in (2.14). And there is: for $\mathrm{N} \geqslant 1$, define ℓ_N as the random variable on the symmetric group $\mathfrak{S}_\mathrm{N}$ that maps a permutation σ to the number of cycles in its canonical cyclic representation (where we count fixed points as cycles of length 1, so, for instance, we have $\ell_\mathrm{N}(1) = \mathrm{N}$). Then, giving $\mathfrak{S}_\mathrm{N}$ the uniform probability measure, we have

$$\mathbf{E}(e^{it\ell_\mathrm{N}}) = \mathrm{N}^{e^{it}-1} \left(\frac{1}{\Gamma(e^{it})} + o(1) \right), \tag{2.15}$$

corresponding to a Poisson distribution with parameter $\log \mathrm{N}$ this time. This is not an isolated property: see the survey paper of Granville [48] for many significant analogies between (multiplicative) properties of integers and random permutations.[2]

Remark 2.4.2 Observe that (2.13) would be true *if* we had a decomposition

$$\omega(n) = \mathrm{Po_N}(n) + \mathrm{Y_N}(n)$$

as random variables on Ω_N, where $\mathrm{Y_N}$ is independent of $\mathrm{Po_N}$ and converges in law to a random variable with characteristic function Φ. However, this is not in fact the case, because Φ is not a characteristic function of a probability measure! (It is unbounded on $\mathbf{R}$.)

Exercise 2.4.3 The goal of this exercise is to give a proof of the formula (2.15). We assume basic familiarity with the notion of tensor product of vector spaces and symmetric powers of vector spaces, and elementary representation theory of finite groups.

For $\mathrm{N} \geqslant 1$, we define ℓ_N as a random variable on $\mathfrak{S}_\mathrm{N}$ as above.

[2] Some readers might also enjoy the comic-book version [49].

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

(1) Show that the formula (2.15) follows from the exact expression

$$\mathbf{E}(e^{it\ell_{\mathrm{N}}}) = \prod_{j=1}^{\mathrm{N}} \left(1 - \frac{1}{j} + \frac{e^{it}}{j}\right)$$

valid for all $\mathrm{N} \geqslant 1$ and all $t \in \mathbf{R}$. [**Hint**: Use the formula

$$\frac{1}{\Gamma(z+1)} = \prod_{k \geqslant 1} \left(1 + \frac{z}{k}\right)\left(1 + \frac{1}{k}\right)^{-z},$$

which is valid for all $z \in \mathbf{C}$ (this is due to Euler).]

(2) Show that (1) is also equivalent with the formula

$$\mathbf{E}(m^{\ell_{\mathrm{N}}}) = \prod_{j=1}^{\mathrm{N}} \left(1 - \frac{1}{j} + \frac{m}{j}\right) \tag{2.16}$$

for all $\mathrm{N} \geqslant 1$ and all *integers* $m \geqslant 0$.

(3) Let $m \geqslant 0$ be a fixed integer. Let V be an m-dimensional complex vector space. For any $\mathrm{N} \geqslant 1$, there is a homomorphism

$$\varrho_{\mathrm{N}} \colon \mathfrak{S}_{\mathrm{N}} \to \mathrm{GL}(\mathrm{V} \otimes \cdots \otimes \mathrm{V}) = \mathrm{GL}(\mathrm{V}^{\otimes \mathrm{N}})$$

(with N tensor factors) such that $\sigma \in \mathfrak{S}_{\mathrm{N}}$ is sent to the unique automorphism of the tensor power $\mathrm{V}^{\otimes \mathrm{N}}$ which satisfies

$$x_1 \otimes \cdots \otimes x_{\mathrm{N}} \mapsto x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(\mathrm{N})}$$

for all $(x_1, \ldots, x_{\mathrm{N}}) \in \mathrm{V}^{\otimes \mathrm{N}}$. (This is a representation of $\mathfrak{S}_{\mathrm{N}}$ on the vector space $\mathrm{V}^{\otimes \mathrm{N}}$; note that this space has dimension m^{N}.)

(4) Show that for any $\sigma \in \mathfrak{S}_{\mathrm{N}}$, the trace of the automorphism $\varrho_{\mathrm{N}}(\sigma)$ of $\mathrm{V}^{\otimes \mathrm{N}}$ is equal to $m^{\ell_{\mathrm{N}}(\sigma)}$.

(5) Deduce that the formula (2.16) holds. [**Hint**: Use the fact that for any representation $\varrho \colon \mathrm{G} \to \mathrm{GL}(\mathrm{E})$ of a finite group on a finite-dimensional $\mathbf{C}$-vector space, the average of the trace of $\varrho(g)$ over $g \in \mathrm{G}$ is equal to the dimension of the space of vectors $x \in \mathrm{E}$ that are invariant, that is, that satisfy $\varrho(g)(x) = x$ for all $g \in \mathrm{G}$ (see, e.g., [70, Prop. 4.3.1] for this); then identify this space to compute its dimension.]

(6) Deduce also from (2.16) that there exists a sequence $(\mathrm{B}_j)_{j \geqslant 1}$ of independent Bernoulli random variables such that we have an equality in law

$$\ell_{\mathrm{N}} = \mathrm{B}_1 + \cdots + \mathrm{B}_{\mathrm{N}}$$

for all $\mathrm{N} \geqslant 1$, and $\mathbf{P}(\mathrm{B}_j = 1) = 1/j$ for all $j \geqslant 1$. (This decomposition is often obtained by what is called the "Chinese Restaurant Process" in the probabilistic literature; see, for instance, [2, Example 2.4].)

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

2.5 Final Remarks

Classically, the Erdős–Wintner and the Erdős–Kac Theorem (and related topics) are presented in a different manner, which is well illustrated in the book of Tenenbaum [115, III.1, III.2]. This emphasizes the notion of *density* of sets of integers, namely, quantities like

$$\limsup_{N\to+\infty} \frac{1}{N} |\{1 \leqslant n \leqslant N \mid n \in A\}|$$

for a given set A, or the associated liminf, or the limit when it exists. Convergence in law is then often encapsulated in the existence of these limits for sets of the form

$$A = \{n \geqslant 1 \mid f(n) \leqslant x\},$$

the limit $F(x)$ (which is only assumed to exist for continuity points of F) being a "distribution function," that is, $F(x) = \mathbf{P}(X \leqslant x)$ for some real-valued random variable X.

Our emphasis on a more systematic probabilistic presentation has the advantage of leading more naturally to the use of purely probabilistic techniques and insights. This will be especially relevant when we consider random variables with values in more complicated sets than $\mathbf{R}$ (as we will do in the next chapters), in which case the analogue of distribution functions becomes awkward or simply doesn't exist. Our point of view is also more natural when we come to consider arithmetic random variables Y_N on Ω_N that genuinely depend on N, in the sense that there doesn't exist an arithmetic function f such that Y_N is the restriction of f to Ω_N for all $N \geqslant 1$.

Among the many generalizations of the Erdős–Kac Theorem (and related results for more general arithmetic functions), we wish to mention Billingsley's work [8, Th. 4.1, Example 1, p. 764] that obtains a functional version where the convergence in law is toward *Brownian motion* (we refer to Billingsley's very accessible text [7] for a first presentation of Brownian motion, and to the book of Revuz and Yor [104] for a complete modern treatment): for $0 \leqslant t \leqslant 1$, define a random variable $\widetilde{X}_N$ on Ω_N with values in the Banach space $C([0,1])$ of continuous functions on $[0,1]$ by putting $\widetilde{X}_N(n)(0) = 0$ and

$$\widetilde{X}_N(n)\left(\frac{\log\log k}{\log\log N}\right) = \frac{1}{(\log\log N)^{1/2}}\left(\sum_{\substack{p|n\\ p\leqslant k}} 1 - \log\log k\right)$$

for $2 \leqslant k \leqslant N$, and by linear interpolation between such points. Then Billingsley proves that $\widetilde{X}_N$ converges in law to Brownian motion as $N \to +\infty$.

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press

Another very interesting limit theorem of Billingsley (see [6] and also [10, Th. I.4.5]) deals with the distribution of *all* the prime divisors of an integer $n \in \Omega_N$, and establishes convergence in law of a suitable normalization of these. Precisely, let X be the compact topological space

$$X = \prod_{k \geqslant 1} [0, 1].$$

For all integers $n \geqslant 1$, denote by

$$p_1 \geqslant p_2 \geqslant \cdots \geqslant p_{\Omega(n)}$$

the prime divisors of n, counted with multiplicity and in nonincreasing order. Moreover, define $p_k = 1$ if $k > \Omega(n)$. Define then an X-valued random variable $\mathsf{D}_N = (\mathsf{D}_{N,k})_{k \geqslant 1}$, where

$$\mathsf{D}_{N,k}(n) = \frac{\log p_k}{\log n}$$

for $n \in \Omega_N$ (in other words, we have $p_k = n^{\mathsf{D}_{N,k}(n)}$). Then Billingsley proved that the random variables D_N converge, as $N \to +\infty$, to a measure on X, which is called the Poisson–Dirichlet distribution (with parameter 1). This measure is quite an interesting one, and occurs also (among other places) in a similar limit theorem for random variables encoding the length of the cycles occurring in a random permutation, again ordered to be nonincreasing (another example of the connections between prime factorizations and permutations which were mentioned in the previous section 2.4).

A shorter proof of this limit theorem was given by Donnelly and Grimmett [27]. It is based on the remark that the Poisson–Dirichlet measure is the image under a certain continuous map of the natural measure on X under which the components of elements of X form a sequence of independent uniformly distributed random variables on $[0, 1]$; arithmetically, it turns out to depend only on the estimate

$$\sum_{p \leqslant N} \frac{\log p}{p} = \log N + O(1),$$

which is at the same level of depth as the Mertens formula (see C.3.1 (3)).

[Further references: Tenenbaum [115].**]**

https://doi.org/10.1017/9781108888226.003 Published online by Cambridge University Press