4

The Distribution of Values of the Riemann Zeta Function, II

Probability tools	Arithmetic tools
Definition of convergence in law (Section B.3)	Riemann zeta function (Section C.4)
Kronecker's Theorem (Th. B.6.5)	Möbius function (Def. C.1.3)
Central Limit Theorem (Th. B.7.2)	Mean-square of $\zeta(s)$ on the critical line (Prop. C.4.1)
Gaussian random variable (Section B.7)	Multiplicative functions (Section C.1)
Lipschitz test functions (Prop. B.4.1)	Euler products (Lemma C.1.4)
Method of moments (Th. B.5.5)	

4.1 Introduction

In this chapter, as indicated previously, we will continue working with the values of the Riemann zeta function, but on the critical line $s = \frac{1}{2} + it$, where the issues are much deeper.

Indeed, the analogue of Theorem 3.1.1 fails for $\tau = 1/2$, which shows that the Riemann zeta function is significantly more complicated on the critical line. However, there is a limit theorem after normalization, due to Selberg, for the logarithm of the Riemann zeta function. To state it, we specify carefully the meaning of $\log \zeta(\frac{1}{2} + it)$. We define a random variable L_{T} on Ω_{T} by putting $\mathsf{L}(t) = 0$ if $\zeta(1/2 + it) = 0$, and otherwise

$$\mathsf{L}_{\mathrm{T}}(t) = \log \zeta\left(\tfrac{1}{2} + it\right),$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

where the logarithm of zeta is the unique branch that is holomorphic on a narrow strip

$$\{s = \sigma + iy \in \mathbf{C} \mid \sigma > \tfrac{1}{2} - \delta, \quad |y - t| \leqslant \delta\}$$

for some $\delta > 0$ and satisfies $\log \zeta(\sigma + it) \to 0$ as $\sigma \to +\infty$.

Theorem 4.1.1 (Selberg) *With notation as above, the random variables*

$$\frac{\mathsf{L}_\mathrm{T}}{\sqrt{\frac{1}{2}\log\log \mathrm{T}}}$$

on Ω_T *converge in law as* $\mathrm{T} \to +\infty$ *to a standard complex Gaussian random variable.*

We will in fact only prove "half" of this theorem: we consider only the real part of $\log \zeta(\frac{1}{2} + it)$, or in other words, we consider $\log|\zeta(\frac{1}{2} + it)|$. So we (re)define the arithmetic random variables L_T on Ω_T by $\mathsf{L}_\mathrm{T}(t) = 0$ if $\zeta(\frac{1}{2} + it) = 0$, and otherwise $\mathsf{L}_\mathrm{T}(t) = \log|\zeta(\frac{1}{2} + it)|$. Note that dealing with the modulus means in particular that we need not worry about the choice of the branch of the logarithm of complex numbers. We will prove:

Theorem 4.1.2 (Selberg) *The random variables*

$$\frac{\mathsf{L}_\mathrm{T}}{\sqrt{\frac{1}{2}\log\log \mathrm{T}}}$$

converge in law as $\mathrm{T} \to +\infty$ *to a standard real Gaussian random variable.*

4.2 Strategy of the Proof of Selberg's Theorem

We present the recent proof of Theorem 4.1.2 due to Radziwiłł and Soundararajan [95]. In comparison with Bagchi's Theorem, the strategy has the common feature of the use of suitable approximations to ζ, and the probabilistic limiting behavior will ultimately derive from the independence and distribution of the vector $t \mapsto (p^{-it})_p$ (as in Proposition 3.2.5). However, one has to be much more careful than in the previous section.

Precisely, the approximation used by Radziwiłł and Soundararajan involves three steps:

- **Step 1**: An approximation of L_T by the random variable $\widetilde{\mathsf{L}}_\mathrm{T}$ given by $t \mapsto \log|\zeta(\sigma_0 + it)|$ for σ_0 sufficiently close to $1/2$ (where σ_0 depends on T).

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

- **Step 2**: For the random variable Z_{T} given by $t \mapsto \zeta(\sigma_0 + it)$, so that $\log|\mathsf{Z}_{\mathrm{T}}| = \widetilde{\mathsf{L}}_{\mathrm{T}}$, an approximation of the *inverse* $1/\mathsf{Z}_{\mathrm{T}}$ by a short Dirichlet polynomial D_{T} of the type

$$\mathsf{D}_{\mathrm{T}}(s) = \sum_{n \geqslant 1} a_{\mathrm{T}}(n)\mu(n)n^{-s},$$

where $a_{\mathrm{T}}(n)$ is zero for n large enough (again, depending on T); here $\mu(n)$ denotes the Möbius function (see Definition C.1.3), and we recall once more that it satisfies

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

if $\mathrm{Re}(s) > 1$ (see Corollary C.1.5). At this point, we get an approximation of L_{T} by $-\log|\mathsf{D}_{\mathrm{T}}|$.
- **Step 3**: An approximation of $|\mathsf{D}_{\mathrm{T}}|$ by what is essentially a short Euler product, namely, by $\exp(-\mathrm{Re}(\mathsf{P}_{\mathrm{T}}))$, where

$$\mathsf{P}_{\mathrm{T}}(t) = \sum_{p^k \leqslant \mathrm{X}} \frac{1}{k} \frac{1}{p^{k(\sigma_0+it)}} \tag{4.1}$$

for suitable X (again depending on T). In this definition, and in all formulas involving such sums below, the condition $p^k \leqslant \mathrm{X}$ is implicitly restricted to integers $k \geqslant 1$. At this point, L_{T} is approximated by $\mathrm{Re}(\mathsf{P}_{\mathrm{T}})$.

Finally, the last probabilistic step is to prove that the random variables

$$\frac{\mathrm{Re}(\mathsf{P}_{\mathrm{T}})}{\sqrt{\frac{1}{2}\log\log \mathrm{T}}}$$

converge in law to a standard Gaussian random variable as $\mathrm{T} \to +\infty$.

None of these steps (except the last) is easy, in comparison with the results discussed up to now, and the specific approximations that are used (namely, the choices of the coefficients $a_{\mathrm{T}}(n)$ as well as of the length parameter X) are quite subtle and by no means obvious (they can be seen to be related to sieve methods). Even the *nature* of the approximation will not be the same in the three steps!

In order to simplify the reading of the proof, we first specify the relevant parameters. We assume from now on that $\mathrm{T} \geqslant e^{e^2}$. We denote by

$$\varrho_{\mathrm{T}} = \sqrt{\tfrac{1}{2}\log\log \mathrm{T}} \geqslant 1$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

the normalizing factor in the theorem. We then define

$$\mathrm{W} = (\log\log\log \mathrm{T})^4 \asymp (\log \varrho_{\mathrm{T}})^4, \qquad \sigma_0 = \frac{1}{2} + \frac{\mathrm{W}}{\log \mathrm{T}} = \frac{1}{2} + \mathrm{O}\left(\frac{(\log \varrho_{\mathrm{T}})^4}{\log \mathrm{T}}\right), \tag{4.2}$$

$$\mathrm{X} = \mathrm{T}^{1/(\log\log\log \mathrm{T})^2} = \mathrm{T}^{1/\sqrt{\mathrm{W}}}. \tag{4.3}$$

Note that we omit the dependency on T in most of these notation. We will also require a further parameter

$$\mathrm{Y} = \mathrm{T}^{1/(\log\log \mathrm{T})^2} = \mathrm{T}^{4/\varrho^4} \leqslant \mathrm{X}. \tag{4.4}$$

We begin by stating the precise approximation statements. All parameters are now fixed as above for the remainder of this chapter. After stating the precise form of each steps of the proof, we will show how they combine to imply Theorem 4.1.2, and finally we will establish these intermediate results.

Proposition 4.2.1 (Moving outside of the critical line) *We have*

$$\mathbf{E}_{\mathrm{T}}\Big(|\mathsf{L}_{\mathrm{T}} - \tilde{\mathsf{L}}_{\mathrm{T}}|\Big) = o(\varrho_{\mathrm{T}})$$

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

We now define properly the Dirichlet polynomials that appear in the second step of the approximation. It is here that the arithmetic subtlety lies, since the definition is quite delicate. We define first

$$m_1 = 100 \log\log \mathrm{T} \asymp \varrho_{\mathrm{T}} \quad \text{and} \quad m_2 = 100 \log\log\log \mathrm{T} \asymp \log \varrho_{\mathrm{T}}. \tag{4.5}$$

We denote by $b_{\mathrm{T}}(n)$ the characteristic function of the set of squarefree integers $n \geqslant 1$ such that all prime factors of n are $\leqslant$ Y, and n has at most m_1 prime factors. We denote by $c_{\mathrm{T}}(n)$ the characteristic function of the set of squarefree integers $n \geqslant 1$ such that all prime factors p of n satisfy $\mathrm{Y} < p \leqslant \mathrm{X}$, and n has at most m_2 prime factors. We associate to these the Dirichlet polynomials

$$\mathrm{B}(s) = \sum_{n \geqslant 1} \mu(n) b_{\mathrm{T}}(n) n^{-s} \quad \text{and} \quad \mathrm{C}(s) = \sum_{n \geqslant 1} \mu(n) c_{\mathrm{T}}(n) n^{-s}$$

for $s \in \mathbf{C}$. Finally, define $\mathrm{D}(s) = \mathrm{B}(s)\mathrm{C}(s)$. The coefficient of n^{-s} in the expansion of $\mathrm{D}(s)$ is the Dirichlet convolution

$$\begin{aligned} \sum_{de=n} b_{\mathrm{T}}(d) c_{\mathrm{T}}(e) \mu(d) \mu(e) &= \sum_{\substack{de=n \\ (d,e)=1}} b_{\mathrm{T}}(d) c_{\mathrm{T}}(e) \mu(d) \mu(e) \\ &= \mu(n) \sum_{\substack{de=n \\ (d,e)=1}} b_{\mathrm{T}}(d) c_{\mathrm{T}}(e) = \mu(n) a_{\mathrm{T}}(n), \end{aligned}$$

say, by Proposition A.4.4, where we used the fact that d and e are coprime if $b_{\mathrm{T}}(d)c_{\mathrm{T}}(e)$ is nonzero since the set of primes dividing an integer in the support of b_{T} is disjoint from the set of primes dividing an integer in the support of c_{T}. It follows then from this formula that $a_{\mathrm{T}}(n)$ is the characteristic function of the set of squarefree integers $n \geqslant 1$ such that

(1) all prime factors of n are $\leqslant \mathrm{X}$;

(2) there are at most m_1 prime factors p of n such that $p \leqslant \mathrm{Y}$;

(3) there are at most m_2 prime factors p of n such that $\mathrm{Y} < p \leqslant \mathrm{X}$.

It is immediate, but very important, that $a_{\mathrm{T}}(n) = 0$ unless n is quite small, namely,

$$n \leqslant \mathrm{Y}^{100 \log\log \mathrm{T}} \mathrm{X}^{100 \log\log\log \mathrm{T}} = \mathrm{T}^c,$$

where

$$c = \frac{100}{\log\log \mathrm{T}} + \frac{100}{\log\log\log \mathrm{T}} \to 0 \quad \text{as } \mathrm{T} \to +\infty.$$

Finally, we define the arithmetic random variable

$$\mathsf{D}_{\mathrm{T}} = \mathrm{D}(\sigma_0 + it). \tag{4.6}$$

Remark 4.2.2 Although the definition of $\mathrm{D}(s)$ may seem complicated, we will see its different components coming together in the proofs of this proposition and the next.

If we consider the support of $a_{\mathrm{T}}(n)$, we note that (by the Erdős–Kac Theorem, restricted to squarefree integers as in Exercise 2.3.4) the typical number of prime factors of an integer $n \leqslant \mathrm{Y}^{m_1}$ is about $\log\log \mathrm{Y}^{m_1} \sim \log\log \mathrm{T}$. Therefore the integers satisfying $b_{\mathrm{T}}(n) = 1$ are quite typical, and only extreme outliers (in terms of the number of prime factors) are excluded. On the other hand, the integers satisfying $c_{\mathrm{T}}(n) = 1$ have much fewer prime factors than is typical, and are therefore quite rare (they are, in a weak sense, "almost prime"). This indicates that a_{T} is a subtle *arithmetic* truncation of the characteristic function of integers $n \leqslant \mathrm{T}^c$, and hence that

$$\sum_{n \geqslant 1} a_{\mathrm{T}}(n)\mu(n)n^{-s}$$

is an arithmetic truncation of the Dirichlet series that formally gives the inverse of $\zeta(s)$. This should be contrasted with the more traditional *analytic* truncations of $\zeta(s)$ that were used in Lemma 3.2.10 and Proposition 3.2.11.

For comparison, it is useful to note that Selberg himself used in many applications certain truncations that are roughly of the shape

$$\sum_{n\leqslant X} \frac{\mu(n)}{n^s}\left(1-\frac{\log n}{\log X}\right).$$

Proposition 4.2.3 (Dirichlet polynomial approximation) *The difference* $\mathsf{Z}_T\mathsf{D}_T$ *converges to* 1 *in* L^2*, that is, we have*

$$\lim_{T\to+\infty} \mathbf{E}_T\big(|1-\mathsf{Z}_T\mathsf{D}_T|^2\big)=0.$$

Proposition 4.2.4 (Euler product approximation) *The random variables* $\mathsf{D}_T\exp(-\mathsf{P}_T)$ *converge to* 1 *in probability, that is, for any* $\varepsilon>0$*, we have*

$$\lim_{T\to+\infty} \mathbf{P}_T\big(|\mathsf{D}_T\exp(\mathsf{P}_T)-1|>\varepsilon\big)=0.$$

In particular, $\mathbf{P}_T(\mathsf{D}_T=0)$ *tends to* 0 *as* $T\to+\infty$.

Despite our probabilistic presentation, the three previous statement are really theorems of number theory, and would usually be stated without probabilistic notation. For instance, Proposition 4.2.1 means that

$$\frac{1}{T}\int_{-T}^{T} |\log|\zeta(1/2+it)|-\log|\zeta(\sigma_0+it)||dt=o(\sqrt{\log\log T}).$$

The last result finally introduces the probabilistic behavior,

Proposition 4.2.5 (Gaussian Euler products) *The random variables* $\varrho_T^{-1}\mathsf{P}_T$ *converge in law as* $T\to+\infty$ *to a standard complex Gaussian random variable. In particular, the random variables*

$$\frac{\mathrm{Re}(\mathsf{P}_T)}{\sqrt{\frac{1}{2}\log\log T}}$$

converge in law to a standard Gaussian random variable.

We will now explain how to combine these ingredients for the final step of the proof.

Proof of Theorem 4.1.2 Until Proposition 4.2.5 is used, this is essentially a variant of the fact that convergence in probability implies convergence in law, and that convergence in L^1 or L^2 implies convergence in probability.

For the details, fix some standard Gaussian random variable $\mathcal{N}$. Let f be a bounded Lipschitz function $\mathbf{R}\longrightarrow\mathbf{R}$, and let $C\geqslant 0$ be a real number such that

$$|f(x)-f(y)|\leqslant C|x-y|,\qquad |f(x)|\leqslant C,\qquad \text{for } x,\ y\in\mathbf{R}.$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

We consider the difference

$$\left|\mathbf{E}_T\left(f\left(\frac{\mathsf{L}_T}{\varrho_T}\right)\right) - \mathbf{E}(f(\mathcal{N}))\right|$$

and must show that this tends to 0 as $T \to +\infty$.

We estimate this quantity using the "chain" of approximations introduced above: we have

$$\begin{aligned}
&\left|\mathbf{E}_T\left(f\left(\frac{\mathsf{L}_T}{\varrho_T}\right)\right) - \mathbf{E}(f(\mathcal{N}))\right| \\
&\quad \leqslant \mathbf{E}_T\left(\left|f\left(\frac{\mathsf{L}_T}{\varrho_T}\right) - f\left(\frac{\widetilde{\mathsf{L}}_T}{\varrho_T}\right)\right|\right) + \mathbf{E}_T\left(\left|f\left(\frac{\widetilde{\mathsf{L}}_T}{\varrho_T}\right) - f\left(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\right)\right|\right) \\
&\qquad + \mathbf{E}_T\left(\left|f\left(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\right) - f\left(\frac{\operatorname{Re}\mathsf{P}_T}{\varrho_T}\right)\right|\right) \\
&\qquad + \left|\mathbf{E}_T\left(f\left(\frac{\operatorname{Re}\mathsf{P}_T}{\varrho_T}\right)\right) - \mathbf{E}(f(\mathcal{N}))\right|,
\end{aligned} \tag{4.7}$$

and we discuss each of the four terms on the right-hand side using the four previous propositions (here and below, we define $|\mathsf{D}_T|^{-1}$ to be 0 if $\mathsf{D}_T = 0$).

The first term is handled straightforwardly using Proposition 4.2.1: we have

$$\mathbf{E}_T\left(\left|f\left(\frac{\mathsf{L}_T}{\varrho_T}\right) - f\left(\frac{\widetilde{\mathsf{L}}_T}{\varrho_T}\right)\right|\right) \leqslant \frac{C}{\varrho_T}\mathbf{E}_T(|\mathsf{L}_T - \widetilde{\mathsf{L}}_T|) \longrightarrow 0$$

as $T \to +\infty$.

For the second term, let $A_T \subset \Omega_T$ be the event

$$\{\mathsf{D}_T = 0\} \cup \{|\widetilde{\mathsf{L}}_T - \log|\mathsf{D}_T|^{-1}| > 1/2\}$$

and A'_T its complement. Since $\log|\mathsf{Z}_T| = \widetilde{\mathsf{L}}_T$, we then have

$$\mathbf{E}_T\left(\left|f\left(\frac{\widetilde{\mathsf{L}}_T}{\varrho_T}\right) - f\left(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\right)\right|\right) \leqslant 2C\,\mathbf{P}_T(A_T) + \frac{C}{2\varrho_T}.$$

Proposition 4.2.3 implies that $\mathbf{P}_T(A_T) \to 0$ (convergence to 1 of $\mathsf{Z}_T\mathsf{D}_T$ in L^2 implies convergence to 1 in probability, hence convergence to 0 in probability for the logarithm of the modulus) and therefore

$$\mathbf{E}_T\left(\left|f\left(\frac{\widetilde{\mathsf{L}}_T}{\varrho_T}\right) - f\left(\frac{\log|\mathsf{D}_T|^{-1}}{\varrho_T}\right)\right|\right) \to 0$$

as $T \to +\infty$.

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

We now come to the third term on the right-hand side of (4.7). Distinguishing according to the events

$$\mathrm{B_T} = \left\{ \left| \log |\mathsf{D}_\mathrm{T} \exp(\mathsf{P}_\mathrm{T})| \right| > 1/2 \right\}$$

and its complement, we get as before

$$\mathbf{E}_\mathrm{T}\left(\left| f\left(\frac{\log |\mathsf{D}_\mathrm{T}|^{-1}}{\varrho_\mathrm{T}} \right) - f\left(\frac{\operatorname{Re} \mathsf{P}_\mathrm{T}}{\varrho_\mathrm{T}} \right) \right| \right) \leqslant 2\mathrm{C}\, \mathbf{P}_\mathrm{T}(\mathrm{B_T}) + \frac{\mathrm{C}}{2\varrho_\mathrm{T}},$$

and this also tends to 0 as $\mathrm{T} \to +\infty$ by Proposition 4.2.4.

Finally, Proposition 4.2.5 implies that

$$\left| \mathbf{E}_\mathrm{T}\left(f\left(\frac{\operatorname{Re} \mathsf{P}_\mathrm{T}}{\varrho_\mathrm{T}} \right) \right) - \mathbf{E}(f(\mathcal{N})) \right| \to 0$$

as $\mathrm{T} \to +\infty$, and hence we conclude the proof of the theorem, assuming the approximation statements. □

We now explain the proofs of these four propositions. We begin with the easiest part, which also happens to be where the transition to the pure probabilistic behavior happens. A key tool is the quantitative form of Proposition 3.2.5 contained in Lemma 3.2.6. More precisely, as in Section 3.2, let $\mathrm{X} = (\mathrm{X}_p)_p$ be a sequence of independent random variables uniformly distributed on $\mathbf{S}^1$. We define X_n for $n \geqslant 1$ by multiplicativity as in formula (3.7).

Lemma 4.2.6 *Let $(a(n))_{n\geqslant 1}$ be any sequence of complex numbers with bounded support. For any $\mathrm{T} \geqslant 2$ and $\sigma \geqslant 0$, we have*

$$\begin{aligned}\mathbf{E}_\mathrm{T}\left(\left| \sum_{n\geqslant 1} \frac{a(n)}{n^{\sigma+it}} \right|^2 \right) &= \sum_{n\geqslant 1} \frac{|a(n)|^2}{n^{2\sigma}} + \mathrm{O}\left(\frac{1}{\mathrm{T}} \sum_{\substack{m,n\geqslant 1\\ m\neq 1}} \frac{|a(m)a(n)|}{(mn)^{\sigma-\frac{1}{2}}} \right)\\ &= \mathbf{E}\left(\left| \sum_{n\geqslant 1} \frac{\mathrm{X}_n}{n^{\sigma}} \right|^2 \right) + \mathrm{O}\left(\frac{1}{\mathrm{T}} \sum_{\substack{m,n\geqslant 1\\ m\neq 1}} \frac{|a(m)a(n)|}{(mn)^{\sigma-\frac{1}{2}}} \right),\end{aligned}$$

where the implied constant is absolute.

Proof We have

$$\mathbf{E}_\mathrm{T}\left(\left| \sum_{n\geqslant 1} \frac{a(n)}{n^{\sigma+it}} \right|^2 \right) = \sum_m \sum_n \frac{a(m)\overline{a(n)}}{(mn)^{\sigma}} \mathbf{E}_\mathrm{T}\left(\left(\frac{n}{m} \right)^{it} \right).$$

We now apply Lemma 3.2.6 and separate the "diagonal" contribution where $m = n$ from the remainder. This leads to the first formula in the lemma, and the second then reflects the orthonormality of the sequence $(\mathrm{X}_n)_{n\geqslant 1}$. □

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

When applying this lemma, we call the first term the "diagonal" contribution and the second the "off-diagonal" one.

Proof of Proposition 4.2.5 We have $\mathsf{P}_\mathrm{T} = \mathsf{Q}_\mathrm{T} + \mathsf{R}_\mathrm{T}$, where Q_T is the contribution of the primes and R_T the contribution of squares and higher powers of primes. We first claim that R_T is uniformly bounded in L^2 for all T. Indeed, using Lemma 4.2.6, we get

$$\begin{aligned}\mathbf{E}_\mathrm{T}(|\mathsf{R}_\mathrm{T}|^2) &= \mathbf{E}_\mathrm{T}\Bigl(\Bigl|\sum_{k\geqslant 2}\sum_{p\leqslant \mathrm{X}^{1/k}} \frac{1}{k} p^{-k\sigma_0} p^{-kit}\Bigr|^2\Bigr)\\ &= \sum_{\substack{p^k\leqslant \mathrm{X}\\ k\geqslant 2}} \frac{1}{k^2} p^{-2k\sigma_0} + \mathrm{O}\Bigl(\frac{1}{\mathrm{T}}\sum_{k,l\geqslant 2}\sum_{\substack{p^k,q^l\leqslant \mathrm{X}\\ p\neq q}} \frac{1}{kl}(pq)^{-2\sigma_0+\frac{1}{2}}\Bigr)\\ &\ll 1 + \frac{\mathrm{X}^2\log \mathrm{X}}{\mathrm{T}} \ll 1\end{aligned}$$

since $\mathrm{X} \ll \mathrm{T}^\varepsilon$ for any $\varepsilon > 0$.

From this, it follows that it is enough to show that $\mathsf{Q}_\mathrm{T}/\varrho_\mathrm{T}$ converges in law to a standard complex Gaussian random variable $\mathcal{N}$. For this purpose, we use moments, that is, we compute

$$\mathbf{E}_\mathrm{T}\Bigl(\mathsf{Q}_\mathrm{T}^k\overline{\mathsf{Q}_\mathrm{T}^\ell}\Bigr)$$

for integers $k, \ell \geqslant 0$, and we compare with the corresponding moment of the random variable

$$\mathrm{Q}_\mathrm{T} = \sum_{p\leqslant \mathrm{X}} p^{-\sigma_0}\mathrm{X}_p.$$

After applying Lemma 3.2.6 again (as in the proof of Lemma 4.2.6), we find that

$$\mathbf{E}_\mathrm{T}\Bigl(\mathsf{Q}_\mathrm{T}^k\overline{\mathsf{Q}_\mathrm{T}^\ell}\Bigr) = \mathbf{E}(\mathrm{Q}_\mathrm{T}^k\overline{\mathrm{Q}}_\mathrm{T}^\ell) + \mathrm{O}\Bigl(\frac{1}{\mathrm{T}}\sum_{m\neq n}(mn)^{-\sigma_0+1/2}\Bigr),$$

where the sum in the error term runs over integers m (resp. n) with at most k prime factors, counted with multiplicity, all of which are $\leqslant$ X (resp. at most ℓ prime factors, counted with multiplicity, all of which are $\leqslant$ X). Hence this error term is

$$\ll \frac{1}{\mathrm{T}}\Bigl(\sum_{p\leqslant \mathrm{X}} 1\Bigr)^{k+\ell} \ll \frac{\mathrm{X}^{k+\ell}}{\mathrm{T}}.$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

Next, we note that

$$\mathbf{V}(\mathrm{Q_T}) = \sum_{p \leqslant \mathrm{X}} p^{-2\sigma_0}\, \mathbf{V}(\mathrm{X}_p^2) = \frac{1}{2} \sum_{p \leqslant \mathrm{X}} p^{-2\sigma_0}.$$

We compute this sum by splitting in two ranges $p \leqslant \mathrm{Y}$ and $\mathrm{Y} < p \leqslant \mathrm{X}$ (recall that σ_0 depends on T). The second sum is

$$\ll \sum_{\mathrm{Y} < p \leqslant \mathrm{X}} \frac{1}{p} = \log\left(\frac{\log \mathrm{X}}{\log \mathrm{Y}}\right) + \mathrm{O}(1) \ll \log\log\log \mathrm{T}$$

by Proposition C.3.1 and (4.2). On the other hand, for $p \leqslant \mathrm{Y} = \mathrm{T}^{1/(\log\log \mathrm{T})^2}$, we have

$$p^{-2\sigma_0} = p^{-1} \exp\left(-2\frac{(\log p)}{(\log \mathrm{T})}\mathrm{W}\right) = p^{-1}\left(1 + \mathrm{O}\left(\frac{\mathrm{W}}{(\log\log \mathrm{T})^2}\right)\right),$$

which, in view of (4.2), implies that $\mathbf{V}(\mathrm{Q_T}) \sim \frac{1}{2}\log\log \mathrm{T} = \varrho_{\mathrm{T}}^2$ as $\mathrm{T} \to +\infty$.

It is finally again a case of the Central Limit Theorem that $\mathrm{Q_T}/\sqrt{\mathbf{V}(\mathrm{Q_T})}$, and hence also $\mathrm{Q_T}/\varrho_{\mathrm{T}}$, converges in law to a standard complex Gaussian random variable, with convergence of the moments (Theorem B.7.2 and Theorem B.5.6 (2), Remark B.5.8), so the conclusion follows from the method of moments since $\mathrm{X}^{k+\ell}/\mathrm{T} \to 0$ as $\mathrm{T} \to +\infty$. □

The other propositions will now be proved in order. Some of the arithmetic results that we will used are only stated in Appendix C (with suitable references).

Proof of Proposition 4.2.1 We appeal to Hadamard's factorization of the Riemann zeta function (Proposition C.4.3) in the form of its corollary, Proposition C.4.4. Let $t \in \Omega_{\mathrm{T}}$ be such that there is no zero of zeta with ordinate t (this only excludes finitely many values of t for a given T). We have

$$\begin{aligned}\log|\zeta(\sigma_0 + it)| - \log|\zeta(\tfrac{1}{2} + it)| &= \mathrm{Re}\left(\int_{1/2}^{\sigma_0} \frac{\zeta'}{\zeta}(\sigma + it)\, d\sigma\right)\\ &= \int_{1/2}^{\sigma_0} \mathrm{Re}\left(\frac{\zeta'}{\zeta}(\sigma + it)\right) d\sigma.\end{aligned}$$

For any σ with $\frac{1}{2} \leqslant \sigma \leqslant \sigma_0$, we have

$$-\frac{\zeta'}{\zeta}(\sigma + it) = \sum_{|t - \mathrm{Im}(\varrho)| < 1} \frac{1}{\sigma + it - \varrho} + \mathrm{O}(\log(2 + |t|)),$$

by Proposition C.4.4, where the sum is over zeros ϱ of $\zeta(s)$, counted with multiplicity, such that $|\sigma + it - \varrho| < 1$.

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

We fix $t_0 \in \Omega_{\mathrm{T}}$ and integrate over t such that $|t - t_0| \leqslant 1$. This leads to

$$\int_{t_0-1}^{t_0+1} \big|\log|\zeta(\sigma_0 + it)| - \log|\zeta(\tfrac{1}{2} + it)|\big| dt$$
$$\leqslant \sum_{|\operatorname{Im}(\varrho)-t_0|\leqslant 1} \int_{t_0-1}^{t_0+1} \int_{\frac{1}{2}}^{\sigma_0} \left|\operatorname{Re}\left(\frac{1}{\sigma + it - \varrho}\right)\right| dt d\sigma.$$

An elementary integral (!) gives

$$\int_{t_0-1}^{t_0+1} \left|\operatorname{Re}\left(\frac{1}{\sigma + it - \varrho}\right)\right| dt \leqslant \int_{\mathbf{R}} \left|\operatorname{Re}\left(\frac{1}{\sigma + it - \varrho}\right)\right| dt$$
$$= \int_{\mathbf{R}} \frac{|\sigma - \beta|}{(\sigma - \beta)^2 + (t - \gamma)^2} dt = \pi$$

for all σ and ϱ. Hence we get

$$\frac{1}{\mathrm{T}} \int_{|t-t_0|\leqslant 1} \big|\log|\zeta(\tfrac{1}{2} + it - \varrho)| - \log|\zeta(\sigma_0 + it - \varrho)|\big| dt \ll (\sigma_0 - \tfrac{1}{2}) \frac{m(t_0)}{\mathrm{T}},$$

where $m(t_0)$ is the number of zeros ϱ such that $|t_0 - \operatorname{Im}(\varrho)| \leqslant 1$. This is $\ll \log(2 + |t_0|)$ by Proposition C.4.4 again. Finally, by summing the bound

$$\frac{1}{\mathrm{T}} \int_{|t-t_0|\leqslant 1} \left|\log\left|\zeta\left(\tfrac{1}{2} + it - \varrho\right)\right| - \log|\zeta(\sigma_0 + it - \varrho)|\right| dt$$
$$\ll \left(\sigma_0 - \tfrac{1}{2}\right) \frac{\log(2 + |t_0|)}{\mathrm{T}}$$

over a partition of Ω_{T} in $\ll \mathrm{T}$ intervals of length 2, we deduce that

$$\mathbf{E}_{\mathrm{T}}(|\mathsf{L}_{\mathrm{T}} - \widetilde{\mathsf{L}}_{\mathrm{T}}|) \ll (\sigma_0 - \tfrac{1}{2}) \log \mathrm{T} = \mathrm{W}.$$

We have $\mathrm{W} = o(\varrho_{\mathrm{T}})$ (by a rather wide margin!), and the proposition follows. □

The last two propositions are more involved, and we present their proofs in separate sections.

4.3 Dirichlet Polynomial Approximation

We will prove Proposition 4.2.3 in this section, that is, we need to prove that

$$\mathbf{E}_{\mathrm{T}}(|1 - \mathsf{Z}_{\mathrm{T}}\mathsf{D}_{\mathrm{T}}|^2),$$

where $\mathsf{Z}_{\mathrm{T}}(t) = \zeta(\sigma_0 + it)$, tends to 0 as $\mathrm{T} \to +\infty$. This is arithmetically the most involved part of the proof.

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

First of all, we use the approximation formula

$$\zeta(\sigma_0 + it) = \sum_{1\leqslant n\leqslant \mathrm{T}} n^{-\sigma_0 - it} + \mathrm{O}\left(\frac{\mathrm{T}^{1-\sigma_0}}{|t|+1} + \mathrm{T}^{-1/2}\right)$$

for $t \in \Omega_{\mathrm{T}}$ (see Proposition C.4.5). Multiplying by D_{T}, we obtain

$$\begin{aligned}\mathbf{E}_{\mathrm{T}}(\mathsf{Z}_{\mathrm{T}}\mathsf{D}_{\mathrm{T}}) = &\sum_{\substack{m\geqslant 1\\ n\leqslant \mathrm{T}}} a_{\mathrm{T}}(m)\mu(m)\,\mathbf{E}_{\mathrm{T}}((mn)^{-\sigma_0})\\ &+ \mathrm{O}\Bigl(\mathrm{T}^{1/2}\sum_{m\geqslant 1} a_{\mathrm{T}}(m)\,\mathbf{E}_{\mathrm{T}}((|t|+1)^{-1}) + \mathrm{T}^{-1/2}\sum_{m\geqslant 1} a_{\mathrm{T}}(m)m^{-\sigma_0}\Bigr).\end{aligned}$$

We recall that $|a_{\mathrm{T}}(n)| \leqslant 1$ for all n, and $a_{\mathrm{T}}(n) = 0$ unless $n \ll \mathrm{T}^{\varepsilon}$, for any $\varepsilon > 0$. Hence, by (3.4), this becomes

$$\begin{aligned}\mathbf{E}_{\mathrm{T}}(\mathsf{Z}_{\mathrm{T}}\mathsf{D}_{\mathrm{T}}) &= 1 + \mathrm{O}\Bigl(\frac{1}{\mathrm{T}}\sum_{\substack{n\leqslant \mathrm{T}\\ m\neq n}} a_{\mathrm{T}}(m)(mn)^{-\sigma_0}(\log mn)\Bigr) + \mathrm{O}(\mathrm{T}^{-1/2+\varepsilon})\\ &= 1 + \mathrm{O}(\mathrm{T}^{-1/2+\varepsilon})\end{aligned}$$

for any $\varepsilon > 0$ (in the diagonal terms, only $m = n = 1$ contributes, and in the off-diagonal terms $mn \neq 1$, we have $\mathbf{E}_{\mathrm{T}}((mn)^{-it}) \ll \mathrm{T}^{-1}\log(mn)$). It follows that it suffices to prove that

$$\lim_{\mathrm{T}\to+\infty} \mathbf{E}_{\mathrm{T}}(|\mathsf{Z}_{\mathrm{T}}\mathsf{D}_{\mathrm{T}}|^2) = 1.$$

We expand the mean-square using the formula for D_{T} and obtain

$$\mathbf{E}_{\mathrm{T}}(|\mathsf{Z}_{\mathrm{T}}\mathsf{D}_{\mathrm{T}}|^2) = \sum_{m,n} \frac{\mu(m)\mu(n)}{(mn)^{\sigma_0}} a_{\mathrm{T}}(m)a_{\mathrm{T}}(n)\,\mathbf{E}_{\mathrm{T}}\left(\left(\frac{m}{n}\right)^{it}|\mathsf{Z}_{\mathrm{T}}|^2\right).$$

Now the asymptotic formula of Proposition C.4.6 translates to a formula for $\mathbf{E}_{\mathrm{T}}\left((m/n)^{it}|\mathsf{Z}_{\mathrm{T}}|^2\right)$, namely,

$$\begin{aligned}\mathbf{E}_{\mathrm{T}}\left(\left(\frac{m}{n}\right)^{it}|\mathsf{Z}_{\mathrm{T}}|^2\right) = \zeta(2\sigma_0)&\left(\frac{(m,n)^2}{mn}\right)^{\sigma_0}\\ &+ \zeta(2-2\sigma_0)\left(\frac{(m,n)^2}{mn}\right)^{1-\sigma_0}\mathbf{E}_{\mathrm{T}}\left(\left(\frac{|t|}{2\pi}\right)^{1-2\sigma_0}\right)\\ &+ \mathrm{O}(\min(m,n)\mathrm{T}^{-\sigma_0+\varepsilon})\end{aligned}$$

for any $\varepsilon > 0$, where the expectation is really the integral

$$\frac{1}{2\mathrm{T}}\int_{-\mathrm{T}}^{\mathrm{T}}\left(\frac{|t|}{2\pi}\right)^{1-2\sigma_0} dt,$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

and we recall that (m,n) denotes here the gcd of m and n.

Using the properties of $a_{\mathrm{T}}(n)$, the error term is easily handled, since it is at most

$$\mathrm{T}^{-\sigma_0+\varepsilon}\sum_{m,n}(mn)^{-\sigma_0}a_{\mathrm{T}}(m)a_{\mathrm{T}}(n)\min(m,n) \leqslant \mathrm{T}^{-\sigma_0+\varepsilon}\Bigl(\sum_m m^{1/2}a_{\mathrm{T}}(m)\Bigr)^2$$
$$\ll \mathrm{T}^{-\sigma_0+2\varepsilon}$$

for any $\varepsilon > 0$. Thus we only need to handle the main terms, which we write as

$$\zeta(2\sigma_0)\mathrm{M}_1 + \zeta(2-2\sigma_0)\,\mathbf{E}_{\mathrm{T}}\left(\left(\frac{|t|}{2\pi}\right)^{1-2\sigma_0}\right)\mathrm{M}_2, \tag{4.8}$$

where

$$\mathrm{M}_1 = \sum_{m,n}\frac{\mu(m)\mu(n)}{(mn)^{2\sigma_0}}a_{\mathrm{T}}(m)a_{\mathrm{T}}(n)(m,n)^{2\sigma_0}$$

and M_2 is the other term. Using the multiplicative structure of a_{T}, the first term factors in turn as $\mathrm{M}_1 = \mathrm{M}_1'\mathrm{M}_1''$, where

$$\mathrm{M}_1' = \sum_{m,n}\frac{\mu(m)\mu(n)}{[m,n]^{2\sigma_0}}b_{\mathrm{T}}(m)b_{\mathrm{T}}(n),$$
$$\mathrm{M}_1'' = \sum_{m,n}\frac{\mu(m)\mu(n)}{[m,n]^{2\sigma_0}}c_{\mathrm{T}}(m)c_{\mathrm{T}}(n).$$

We compare M_1' to the similar sum $\tilde{\mathrm{M}}_1'$ where $b_{\mathrm{T}}(n)$ and $b_{\mathrm{T}}(m)$ are replaced by characteristic functions of integers with all prime factors $\leqslant \mathrm{Y}$, forgetting only the requirement to have $\leqslant m_1$ prime factors. By Example C.1.7, we have

$$\tilde{\mathrm{M}}_1' = \prod_{p\leqslant \mathrm{Y}}\left(1-\frac{1}{p^{2\sigma_0}}\right).$$

The difference $\mathrm{M}_1' - \tilde{\mathrm{M}}_1'$ can be bounded from above by

$$2e^{-m_1}\sum_{m,n}\frac{|\mu(m)\mu(n)|}{[m,n]^{2\sigma_0}}e^{\Omega(m)},$$

where the sum runs over integers with all prime factors $\leqslant \mathrm{Y}$ (this step is a case of what is called “Rankin’s trick”: the condition $\Omega(m) > m_1$ is handled by bounding its characteristic function by the nonnegative function $e^{\Omega(m)-m_1}$). Again from Example C.1.7, this is at most

$$2(\log \mathrm{T})^{-100}\prod_{p\leqslant \mathrm{Y}}\left(1+\frac{1+2e}{p}\right) \ll (\log \mathrm{T})^{-90}$$

(by Proposition C.3.6). Thus

$$\mathrm{M}_1' \sim \prod_{p\leqslant \mathrm{Y}} (1 - p^{-2\sigma_0})$$

as $\mathrm{T} \to +\infty$. One deals similarly with the second term M_1'', which turns out to satisfy

$$\mathrm{M}_1'' \sim \prod_{\mathrm{Y}<p\leqslant \mathrm{X}} (1 - p^{-2\sigma_0}),$$

and hence

$$\mathrm{M}_1 \sim \zeta(2\sigma_0) \prod_{p\leqslant \mathrm{X}} (1 - p^{-2\sigma_0}) = \prod_{p>\mathrm{X}} (1 - p^{-2\sigma_0}).$$

Now, by the choice of the parameters, we obtain from the Prime Number Theorem (Theorem C.3.3) the upper bound

$$\sum_{p>\mathrm{X}} p^{-2\sigma_0} \ll \int_{\mathrm{X}}^{+\infty} \frac{1}{t^{2\sigma_0}} \frac{dt}{\log t} \ll \frac{\mathrm{X}^{1-2\sigma_0}}{(2\sigma_0 - 1)\log \mathrm{X}} = \frac{\mathrm{X}^{1-2\sigma_0}}{2\sqrt{\mathrm{W}}} \leqslant \frac{1}{2\sqrt{\mathrm{W}}}.$$

Since this tends to 0 as $\mathrm{T} \to +\infty$, it follows that

$$\begin{aligned}\prod_{p>\mathrm{X}} (1 - p^{-2\sigma_0}) &= \exp\Big(\sum_{p>\mathrm{X}} \Big(\frac{1}{p^{2\sigma_0}} + \mathrm{O}\Big(\frac{1}{p^{4\sigma_0}}\Big)\Big)\Big) \\ &= \exp\Big(-\sum_{p>\mathrm{X}} p^{-2\sigma_0}\Big)(1 + o(1))\end{aligned}$$

converges to 1 as $\mathrm{T} \to +\infty$.

There only remains to check that the second part M_2 of the main term (4.8 tends to 0 as $\mathrm{T} \to +\infty$. We have

$$\begin{aligned}\mathrm{M}_2 &= \sum_{m,n} \frac{\mu(m)\mu(n)}{mn} a_{\mathrm{T}}(m) a_{\mathrm{T}}(n) (m,n)^{2-2\sigma_0} \\ &= \sum_{m,n} \frac{\mu(m)\mu(n)}{[m,n]^{2-2\sigma_0}} a_{\mathrm{T}}(m) a_{\mathrm{T}}(n) (mn)^{1-2\sigma_0}.\end{aligned}$$

The procedure is very similar: we factor $\mathrm{M}_2 = \mathrm{M}_2'\mathrm{M}_2''$, where M_2' has coefficients b_{T} instead of a_{T}, and M_2'' has c_{T}. Applying Example C.1.7 and Rankin's trick to both factors now leads to

$$\begin{aligned}\mathrm{M}_2 &\sim \prod_{p\leqslant \mathrm{X}} \Big(1 + \frac{1}{p^{2-2\sigma_0}} \Big(-\frac{1}{p^{2\sigma_0-1}} - \frac{1}{p^{2\sigma_0-1}} + \frac{1}{p^{4\sigma_0-2}}\Big)\Big) \\ &= \prod_{p\leqslant \mathrm{X}} \Big(1 - \frac{2}{p} + \frac{1}{p^{2\sigma_0}}\Big).\end{aligned}$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

We deduce from this that the contribution of M_2 to (4.8) is

$$\sim \zeta(2-2\sigma_0)\,\mathbf{E}_T\left(\left(\frac{|t|}{2\pi}\right)^{1-2\sigma_0}\right)\prod_{p\leqslant X}\left(1-\frac{2}{p}+\frac{1}{p^{2\sigma_0}}\right).$$

Since $\zeta(s)$ has a pole at $s=1$ with residue 1, this last expression is

$$\ll \frac{T^{1-2\sigma_0}}{2\sigma_0-1}\prod_{p\leqslant X}\left(1-\frac{1}{p}\right) \ll \frac{T^{1-2\sigma_0}}{(2\sigma_0-1)\log X}.$$

In terms of the parameter W, since $2\sigma_0 - 1 = 2W/\log T$ and $X = T^{1/\sqrt{W}}$, the right-hand side is simply $\exp(-2W)W^{-1/2}$, and hence tends to 0 as $T\to+\infty$. This concludes the proof.

4.4 Euler Product Approximation

This section is devoted to the proof of Proposition 4.2.4. We need to prove that $\mathsf{D}_T\exp(\mathsf{P}_T)$ converges to 1 in probability. This involves some extra decomposition of P_T: we write

$$\mathsf{P}_T = \mathsf{Q}_T + \mathsf{R}_T,$$

where Q_T is the contribution to (4.1) of the prime powers $p^k \leqslant Y$.

In addition, for any integer $m \geqslant 0$, we denote by $\exp_m$ the Taylor polynomial of degree m of the exponential function at 0, that is,

$$\exp_m(z) = \sum_{j=0}^{m}\frac{z^j}{j!}.$$

We have an elementary lemma:

Lemma 4.4.1 *Let $z\in\mathbf{C}$ and $m\geqslant 0$. If $m\geqslant 100|z|$, then*

$$\exp_m(z) = e^z + O(\exp(-m)) = e^z(1+O(\exp(-99|z|))).$$

Proof Indeed, since $j!\geqslant (j/e)^k$ for all $j\geqslant 0$ and $|z|\leqslant m/100$, the difference $e^z - \exp_m(z)$ is at most

$$\sum_{j>m}\frac{(m/100)^j}{j!} \leqslant \sum_{j>m}\left(\frac{em}{100j}\right)^j \ll \exp(-m).$$ □

We define

$$\mathsf{E}_T = \exp_{m_1}(-\mathsf{Q}_T) \quad\text{and}\quad \mathsf{F}_T = \exp_{m_2}(-\mathsf{R}_T),$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

where we recall that m_1 and m_2 are the parameters defined in (4.5). We have by definition $\mathsf{D}_\mathrm{T} = \mathsf{B}_\mathrm{T}\mathsf{C}_\mathrm{T}$, with

$$\mathsf{B}_\mathrm{T}(t) = \sum_{n\geqslant 1} b_\mathrm{T}(n)\mu(n)n^{-\sigma_0-it} \quad \text{and} \quad \mathsf{C}_\mathrm{T}(t) = \sum_{n\geqslant 1} c_\mathrm{T}(n)\mu(n)n^{-\sigma_0-it},$$

where b_T and c_T are defined after the statement of Proposition 4.2.1, for example, $b_\mathrm{T}(n)$ is the characteristic function of squarefree integers n such that n has $\leqslant m_1$ prime factors, all of which are $\leqslant \mathrm{Y}$.

The idea of the proof is that, usually, Q_T (resp. R_T) is not too large, and then the random variable E_T is a good approximation to $\exp(-\mathsf{Q}_\mathrm{T})$. On the other hand, because of the shape of E_T (and the choice of the parameters), it will be possible to prove that E_T is close to B_T in L^2-norm, and similarly for F_T and C_T. Combining these facts will lead to the conclusion.

We first observe that, as in the beginning of the proof of Proposition 4.2.5, by the usual appeal to Lemma 3.2.6, we have

$$\mathbf{E}_\mathrm{T}(|\mathsf{Q}_\mathrm{T}|^2) \ll \varrho_\mathrm{T} \quad \text{and} \quad \mathbf{E}_\mathrm{T}(|\mathsf{R}_\mathrm{T}|^2) \ll \log \varrho_\mathrm{T}.$$

Markov's inequality implies that $\mathbf{P}_\mathrm{T}(|\mathsf{Q}_\mathrm{T}| > \varrho_\mathrm{T})$ tends to 0 as $\mathrm{T} \to +\infty$. Now by Lemma 4.4.1, whenever $|\mathsf{Q}_\mathrm{T}| \leqslant \varrho_\mathrm{T}$, we have

$$\mathsf{E}_\mathrm{T} = \exp(-\mathsf{Q}_\mathrm{T})\big(1 + \mathrm{O}((\log \mathrm{T})^{-99})\big).$$

Similarly, the probability $\mathbf{P}_\mathrm{T}(|\mathsf{R}_\mathrm{T}| > \log \varrho_\mathrm{T})$ tends to 0, and whenever $|\mathsf{R}_\mathrm{T}| \leqslant \log \varrho_\mathrm{T}$, we have

$$\mathsf{F}_\mathrm{T} = \exp(-\mathsf{R}_\mathrm{T})\big(1 + \mathrm{O}((\log\log \mathrm{T})^{-99})\big).$$

For the next step, we claim that

$$\mathbf{E}_\mathrm{T}\left(|\mathsf{E}_\mathrm{T} - \mathsf{B}_\mathrm{T}|^2\right) \ll (\log \mathrm{T})^{-60}, \tag{4.9}$$

$$\mathbf{E}_\mathrm{T}\left(|\mathsf{F}_\mathrm{T} - \mathsf{C}_\mathrm{T}|^2\right) \ll (\log\log \mathrm{T})^{-60}. \tag{4.10}$$

We begin the proof of the first estimate with a lemma.

Lemma 4.4.2 *For $t \in \Omega_\mathrm{T}$, we have*

$$\mathsf{E}_\mathrm{T}(t) = \sum_{n\geqslant 1} \alpha(n)n^{-\sigma_0+it},$$

where the coefficients $\alpha(n)$ are zero unless $n \leqslant \mathrm{Y}^{m_1}$ and n has only prime factors $\leqslant \mathrm{Y}$. Moreover $|\alpha(n)| \leqslant 1$ for all n, and $\alpha(n) = \mu(n)b_\mathrm{T}(n)$ if n has $\leqslant m_1$ prime factors, counted with multiplicity, and if there is no prime power p^k dividing n such that $p^k > \mathrm{Y}$.

Proof Since

$$\mathsf{E}_{\mathrm{T}} = \exp_{m_1}(-\mathsf{Q}_{\mathrm{T}}) = \sum_{j=0}^{m_1} \frac{(-1)^j}{j!} \Big(\sum_{p^k \leqslant \mathrm{Y}} \frac{1}{kp^{k(\sigma_0+it)}} \Big)^j,$$

we obtain by expanding the jth power an expression of the desired kind, with coefficients

$$\alpha(n) = \sum_{0 \leqslant j \leqslant m_1} \frac{(-1)^j}{j!} \sum_{\substack{p_1^{k_1} \cdots p_j^{k_j} = n \\ p_i^{k_i} \leqslant \mathrm{Y}}} \frac{1}{k_1 \cdots k_j}.$$

We see from this expression that $\alpha(n)$ is 0 unless $n \leqslant \mathrm{Y}^{m_1}$ and n has only prime factors $\leqslant \mathrm{Y}$. Suppose now that n has $\leqslant m_1$ prime factors, counted with multiplicity, and that no prime power $p^k > \mathrm{Y}$ divides n. Then we may extend the sum defining $\alpha(n)$ to all $j \geqslant 0$, and remove the redundant conditions $p_i^{k_i} \leqslant \mathrm{Y}$, so that

$$\alpha(n) = \sum_{j \geqslant 0} \frac{(-1)^j}{j!} \sum_{p_1^{k_1} \cdots p_j^{k_j} = n} \frac{1}{k_1 \cdots k_j}.$$

But we recognize that this is the coefficient of n^{-s} in the expansion of

$$\exp\Big(-\sum_{k \geqslant 1} \frac{1}{k} p^{-ks} \Big) = \exp(-\log \zeta(s)) = \frac{1}{\zeta(s)} = \sum_{n \geqslant 1} \frac{\mu(n)}{n^s}$$

(viewed as a formal Dirichlet series, or by restricting to $\mathrm{Re}(s) > 1$). This means that, for such integers n, we have $\alpha(n) = \mu(n) = \mu(n) b_{\mathrm{T}}(n)$.

Finally, for any $n \geqslant 1$ now, we have

$$|\alpha(n)| \leqslant \sum_{j \geqslant 0} \frac{1}{j!} \sum_{p_1^{k_1} \cdots p_j^{k_j} = n} \frac{1}{k_1 \cdots k_j} = 1,$$

since the right-hand side is the coefficient of n^{-s} in $\exp(\log \zeta(s)) = \zeta(s)$. ☐

Now define $\delta(n) = \alpha(n) - \mu(n) b_{\mathrm{T}}(n)$ for all $n \geqslant 1$. We have

$$\mathbf{E}_{\mathrm{T}}\left(|\mathsf{E}_{\mathrm{T}} - \mathsf{B}_{\mathrm{T}}|^2\right) = \mathbf{E}_{\mathrm{T}}\bigg(\Big| \sum_{n \geqslant 1} \frac{\delta(n)}{n^{\sigma_0+it}} \Big|^2 \bigg),$$

which we estimate using Lemma 4.2.6. The contribution of the off-diagonal term is

$$\ll \frac{1}{\mathrm{T}} \sum_{m,n \leqslant \mathrm{Y}^{m_1}} |\delta(n)\delta(m)|(mn)^{\frac{1}{2}-\sigma_0} \leqslant \frac{4}{\mathrm{T}} \Big(\sum_{m \leqslant \mathrm{Y}^{m_1}} 1 \Big)^2 \ll \mathrm{T}^{-1+\varepsilon}$$

for any $\varepsilon > 0$, hence is negligible. The diagonal term is

$$\mathrm{M} = \sum_{n \geqslant 1} \frac{|\delta(n)|^2}{n^{2\sigma_0}} \leqslant \sum_{n \geqslant 1} \frac{|\delta(n)|^2}{n}.$$

By Lemma 4.4.2, we have $\delta(n) = 0$ unless either n has $> m_1$ prime factors, counted with multiplicity, or is divisible by a power p^k such that $p^k > \mathrm{Y}$ (and necessarily $p \leqslant \mathrm{Y}$ since δ is supported on integers only divisible by such primes). The contribution of the integers satisfying the first property is at most

$$\sum_{\substack{\Omega(n) > m_1 \\ p \mid n \Rightarrow p \leqslant \mathrm{Y}}} \frac{1}{n}.$$

We use Rankin's trick once more to bound this from above: for any fixed real number $\eta > 1$, we have

$$\sum_{\substack{\Omega(n) > m_1 \\ p \mid n \Rightarrow p \leqslant \mathrm{Y}}} \frac{1}{n} \leqslant \eta^{-m_1} \prod_{p \leqslant \mathrm{Y}} \Big(1 + \frac{\eta}{p} + \cdots \Big) \ll \eta^{-m_1} (\log \mathrm{Y})^{\eta}$$
$$\leqslant (\log \mathrm{T})^{-100 \log \eta + \eta}$$

by Proposition C.3.6. Selecting $\eta = e^{2/3} \leqslant 2$, for instance, this shows that this contribution is $\ll (\log \mathrm{T})^{-60}$.

The contribution of integers divisible by $p^k > \mathrm{Y}$ is at most

$$\Big(\sum_{\substack{p \leqslant \mathrm{Y} \\ p^k > \mathrm{Y}}} \frac{1}{p^k} \Big) \Big(\sum_{p \mid n \Rightarrow p \leqslant \mathrm{Y}} \frac{1}{n} \Big) \leqslant \frac{1}{\mathrm{Y}} \Big(\sum_{\substack{\sqrt{\mathrm{Y}} < p^k \leqslant \mathrm{Y} \\ k \geqslant 2}} 1 \Big) \prod_{p \leqslant \mathrm{Y}} \frac{1}{1 - p^{-1}}$$
$$\ll \mathrm{Y}^{-1/2} (\log \mathrm{Y}),$$

which is even smaller. This concludes the proof of (4.9).

The proof of the second estimate (4.10) is quite similar, with one extra consideration to handle. Indeed, arguing as in Lemma 4.4.2, we obtain the expression

$$\mathsf{F}_{\mathrm{T}}(t) = \sum_{n \geqslant 1} \beta(n) n^{-\sigma_0 + it},$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

for $t \in \Omega_T$, where the coefficients $\beta(n)$ are zero unless $n \leqslant X^{m_2}$ and n has only prime factors $\leqslant X$, satisfy $|\beta(n)| \leqslant 1$ for all n, and finally satisfy $\beta(n) = \mu(n)$ if n has $\leqslant m_2$ prime factors, counted with multiplicity, and if there is no prime power p^k dividing n with $Y < p^k \leqslant X$.

Using this, and defining now $\delta(n) = \beta(n) - \mu(n)c_T(n)$, we get from Lemma 4.2.6 the bound

$$\mathbf{E}_T\left(|\mathsf{F}_T - \mathsf{C}_T|^2\right) \ll \sum_{\substack{n \geqslant 1 \\ \delta(n) \neq 0}} \frac{1}{n^{2\sigma_0}} \leqslant \sum_{\substack{n \geqslant 1 \\ \delta(n) \neq 0}} \frac{1}{n}.$$

But the integers that satisfy $\delta(n) \neq 0$ must be of one of the following types:

(1) Those with $c_T(n) = 1$, which (by the previous discussion) must either have $\Omega(n) > m_2$ (and be divisible by primes $\leqslant X$ only), *or* must be divisible by a prime power p^k such that $p^k > X$ (the possibility that $p^k \leqslant Y$ is here excluded, because $c_T(n) = 1$ implies that n has no prime factor $< Y$). The contribution of these integers is handled as in the case of the bound (4.9) and is $\ll (\log\log T)^{-60}$.

(2) Those with $c_T(n) = 0$ and $\beta(n) \neq 0$; since

$$\beta(n) = \sum_{0 \leqslant j \leqslant m_2} \frac{(-1)^j}{j!} \sum_{\substack{p_1^{k_1} \cdots p_j^{k_j} = n \\ Y < p_i^{k_i} \leqslant X}} \frac{1}{k_1 \cdots k_j},$$

as in the beginning of the proof of Lemma 4.4.2, such an integer n has at least one factorization $n = p_1^{k_1} \cdots p_j^{k_j}$ for some $j \leqslant m_2$, where each prime power $p_i^{k_i}$ is between Y and X. Since $c_T(n) = 0$, either $\Omega(n) > m_2$, *or* n has a prime factor $p > X$, *or* n has a prime factor $p \leqslant Y$. The first two possibilities are again handled exactly like in the proof of (4.9), but the third is somewhat different. We proceed as follows to estimate its contribution, say, N. We have

$$N = \sum_{0 \leqslant j < m_2} N_j,$$

where

$$N_j = \sum_{p \leqslant Y} \sum_{\substack{n = p^k p_1^{k_1} \cdots p_j^{k_j} \\ Y < p_i^{k_i} \leqslant X}} \frac{1}{n}$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

is the contribution of integers with a factorization of length $j+1$ as a product of prime powers between Y and X. By multiplicativity, we get

$$\mathrm{N}_j \leqslant \Big(\sum_{p\leqslant \mathrm{Y}}\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k}\Big)\Big(\sum_{p\leqslant \mathrm{X}}\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k}\Big)^{j-1}.$$

Consider the first factor. For a given prime $p\leqslant \mathrm{Y}$, let l be the smallest integer such that $p^l>\mathrm{Y}$. The sum over k is then

$$\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k}\leqslant \frac{1}{p^l}+\frac{1}{p^{l+1}}+\cdots \ll \frac{1}{p^l}\leqslant \frac{1}{\mathrm{Y}},$$

so that the first factor is $\ll \pi(\mathrm{Y})/\mathrm{Y}\ll(\log \mathrm{Y})^{-1}$. On the other hand, for the second factor, we have

$$\begin{aligned}\sum_{p\leqslant \mathrm{X}}\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k} &= \sum_{p\leqslant \mathrm{Y}}\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k}+\sum_{\mathrm{Y}<p\leqslant \mathrm{X}}\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k}\\ &\ll \frac{\pi(\mathrm{Y})}{\mathrm{Y}}+\sum_{\mathrm{Y}<p\leqslant \mathrm{X}}\sum_{\mathrm{Y}<p^k\leqslant \mathrm{X}}\frac{1}{p^k},\end{aligned}$$

where we used the bound arising from the first factor. For a given prime p with $\mathrm{Y}<p\leqslant \mathrm{X}$, the last sum over k is

$$\frac{1}{p}+\frac{1}{p^2}+\cdots \ll \frac{1}{p},$$

and the sum over p is therefore

$$\sum_{\mathrm{Y}<p\leqslant \mathrm{X}}\frac{1}{p}=\log\Big(\frac{\log \mathrm{X}}{\log \mathrm{Y}}\Big)+\mathrm{O}(1)=\log\log\log \mathrm{T}+\mathrm{O}(1),$$

using the values of X and Y and Proposition C.3.1. Hence the final estimate is

$$\mathrm{N}\ll \frac{1}{\log \mathrm{Y}}(\log\log\log \mathrm{T})^{m_2}\ll(\log\log \mathrm{T})(\log\log\log \mathrm{T})^{m_2}(\log \mathrm{T})^{-1}\to 0$$

as $\mathrm{T}\to+\infty$, from which we finally deduce that (4.10) holds.

With the mean-square estimates (4.9) and (4.10) in hand, we can now finish the proof of Proposition 4.2.5. Except on sets of measure tending to 0 as $\mathrm{T}\to+\infty$, we have

$$\mathsf{B}_\mathrm{T}=\mathsf{E}_\mathrm{T}+\mathrm{O}((\log \mathrm{T})^{-25}),\qquad \mathsf{E}_\mathrm{T}=\exp(-\mathsf{Q}_\mathrm{T})\big(1+\mathrm{O}((\log \mathrm{T})^{-99})\big),$$
$$\frac{1}{\log \mathrm{T}}\leqslant \exp(-\mathsf{Q}_\mathrm{T})\leqslant(\log \mathrm{T})$$

(where the first property follows from (4.9)), and hence

$$\mathsf{B}_\mathrm{T}=\exp(-\mathsf{Q}_\mathrm{T})\big(1+\mathrm{O}((\log \mathrm{T})^{-20})\big),$$

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

again outside of a set of measure tending to 0. Similarly, using (4.10), we get

$$\mathsf{C}_{\mathrm{T}} = \exp(-\mathsf{R}_{\mathrm{T}})\big(1 + \mathrm{O}((\log\log \mathrm{T})^{-20})\big)$$

outside of a set of measure tending to 0. Multiplying the two equalities shows that

$$\mathsf{D}_{\mathrm{T}} = \exp(-\mathsf{P}_{\mathrm{T}})\big(1 + \mathrm{O}((\log\log \mathrm{T})^{-20})\big)$$

with probability tending to 1 as $\mathrm{T} \to +\infty$. This concludes the proof.

Exercise 4.4.3 Try to see what happens if one uses a single range $p^k \leqslant \mathrm{X}$, instead of having the distinction between $p^k \leqslant \mathrm{Y}$ and $\mathrm{Y} < p^k \leqslant \mathrm{X}$.

4.5 Further Topics

Generalizations of Selberg's Central Limit Theorem are much harder to come by than those of Bagchi's Theorem (which is another illustration of the fact that arithmetic L-functions have much more delicate properties on the critical line). There are very few other cases than that of the Riemann zeta function where such a statement is known (see the remarks in [95, §7] for references). For instance, consider the family of modular forms f that is described in Section 3.4. The natural question is now to consider the distribution (possibly with weights ω_f) of $\mathrm{L}(f, \frac{1}{2})$. First, it is a known fact (due to Waldspurger and Kohnen–Zagier) that $\mathrm{L}(f, \frac{1}{2}) \geqslant 0$ in that case. This property reflects a different type of expected distribution of the values $\mathrm{L}(f, \frac{1}{2})$, namely, one expects that the correct normalization is

$$f \mapsto \frac{\log \mathrm{L}(f, \frac{1}{2}) + \frac{1}{2}\log\log q}{\sqrt{\log\log q}},$$

in the sense that this defines a sequence of random variables on Ω_q that should converge in law to a standard (real) Gaussian random variable. Now observe that such a statement, if true, would immediately imply that the proportion of $f \in \Omega_q$ with $\mathrm{L}(f, \frac{1}{2}) = 0$ tends to 0 as $q \to +\infty$, and this is not currently known (this would indeed be a major result in the analytic theory of modular forms).

Nevertheless, there has been significant progress in this direction, for various families, in recent and ongoing work of Radziwiłł and Soundararajan. In [96], they prove *sub-Gaussian* upper bounds for the distribution of L-values in certain families similar to Ω_q (specifically, quadratic twists of a fixed modular form). In [97], they announce Gaussian lower bounds, but for families conditioned to have $\mathrm{L}(f, \frac{1}{2}) \neq 0$ (which, for a number of cases, is known to be a subfamily with positive density as the size tends to infinity).

https://doi.org/10.1017/9781108888226.005 Published online by Cambridge University Press

In addition to these developments, it should be emphasized that Selberg's Theorem serves as a general guiding principle when studying any probabilistic question for the Riemann zeta function on the critical line, and the ideas in its proof are often the starting points toward other results. Indeed, some of the deepest works in probabilistic number theory in recent years have been devoted to studies of finer aspects of the distribution of the Riemann Zeta function on the critical line. A particular focus has been a conjecture of Fedorov, Hiary and Keating [39] that addresses the distribution of the maximum of $\zeta(1/2+it)$ when t varies over an interval of length 1 (and t is taken uniformly at random in $[-\mathrm{T},\mathrm{T}]$ or $[\mathrm{T},2\mathrm{T}]$ with $\mathrm{T}\to+\infty$). This leads to links with objects like log-correlated fields, branching random walks, or Gaussian multiplicative chaos. We refer to the Bourbaki seminar survey of Harper [54] for a discussion of the work of Najnudel [90] and Arguin–Belius–Bourgade–Radziwi–Soundararajan [1], and to Harper's recent preprint [55] for the latest developments in this direction.

One of the reasons that Central Limit Theorems are expected to hold is that they are known to follow from the widely believed moment conjectures for families of L-functions, which predict (with considerable evidence, theoretic, numerical and heuristic) the asymptotic behavior of the Laplace or Fourier transform of the logarithm of the special values of the L-functions. In other words, taking the example of the Riemann zeta function, these conjectures (due to Keating and Snaith [63]) predict the asymptotic behavior of

$$\mathbf{E}_{\mathrm{T}}(e^{s\log|\zeta(\frac{1}{2}+it)|})=\mathbf{E}_{\mathrm{T}}(|\zeta(\tfrac{1}{2}+it)|^{s})=\frac{1}{2\mathrm{T}}\int_{-\mathrm{T}}^{\mathrm{T}}|\zeta(\tfrac{1}{2}+it)|^{s}dt$$

for suitable $s\in\mathbf{C}$. It is of considerable interest that, besides natural arithmetic factors (related to the independence of Proposition 3.2.5 or suitable analogues), these conjectures involve certain terms which originate in Random Matrix Theory. In addition to implying straightforwardly the Central Limit Theorem, note that the moment conjectures also immediately yield the generalization of (3.11) or (3.15), hence can be allowed to deduce general versions of Bagchi's Theorem and universality. Moreover, these moment conjectures (in suitably uniform versions) are also able to settle other interesting conjectures concerning the distribution of values of $\zeta(\frac{1}{2}+it)$. For instance, as shown by Kowalski and Nikeghbali [78], they are known to imply that the image of $t\mapsto\zeta(\frac{1}{2}+it)$, for $t\in\mathbf{R}$, is *dense* in $\mathbf{C}$ (a conjecture of Ramachandra).

[Further references: Katz–Sarnak [62], Blomer, Fouvry, Kowalski, Michel, Milićević, and Sawin [11].**]**